

Exploring the influence of contextual factors of the clinical encounter on clinical reasoning success : (unraveling context specificity)

Citation for published version (APA):

Durning, S. J. (2011). *Exploring the influence of contextual factors of the clinical encounter on clinical reasoning success : (unraveling context specificity)*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20110909sd>

Document status and date:

Published: 01/01/2011

DOI:

[10.26481/dis.20110909sd](https://doi.org/10.26481/dis.20110909sd)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Exploring the Influence of Contextual Factors of the Clinical Encounter on Clinical Reasoning Success

(Unraveling context specificity)

The research reported here was carried out at



in the School of Health Professions Education.



This research project was partially funded by Uniformed Services University of the Health Sciences start-up grant RO83VC.



ISBN 978 94 6159 069 5

Production Datawyse bv / Universitaire Pers Maastricht

Exploring the Influence of Contextual Factors of the Clinical Encounter on Clinical Reasoning Success

(Unraveling context specificity)

Dissertation

to obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus Prof.dr. G. P. M. F. Mols,
in accordance with the decision of the Board of Deans,
to be defended in public on Friday, 9 September 2011 at 12.00 hours

by

Steven J. Durning



Supervisor(s)

Prof.dr. L.W.T. Schuwirth
Prof.dr. C.P.M. van der Vleuten

Co-supervisor

Prof.dr. L. N. Pangaro, Uniformed Services University of the Health Sciences

Assessment Committee

Prof.dr. J.J.G. van Merriënboer (voorzitter)
Prof.dr. Th.J. ten Cate, UMC Utrecht
Prof.dr. R.P. Koopmans
Prof.dr. H.G. Schmidt, EU Rotterdam
Dr. M. Van de Wiel

Table of Contents

Chapter 1	Introduction	7
Chapter 2	Redefining Context in the Clinical Encounter: Implications for Research and Training in Medical Education Published in <i>Academic Medicine</i> 2010;85:894–901	19
Chapter 3	Aging and Cognitive Performance: Challenges and Implications for Physicians Practicing in the 21st Century Published in <i>Journal of Continuing Education in the Health Professions</i> 2010;30:1–8	35
Chapter 4	Making Use of Contrasting Participant Views of the Same Encounter Published in <i>Medical Education</i> 2010;44:953–961	51
Chapter 5	The Feasibility, Reliability, and Validity of a Post-encounter Form for Evaluating Clinical Reasoning Accepted for publication, <i>Medical Teacher</i> , 2011	65
Chapter 6	The Impact of Selected Contextual Factors on Internal Medicine Experts' Diagnostic and Therapeutic Reasoning Accepted for publication, <i>Advances in Health Science Education</i> , 2011	77
Chapter 7	Context and Clinical Reasoning: Understanding the Situation from the Perspective of the Expert's Voice Accepted for publication, <i>Medical Education</i> , 2011	93
Chapter 8	The Effect of Increasing Authenticity of Instructional Format on Student Performance: A Prospective Randomized Trial Accepted for publication, <i>Medical Education</i> , 2011	109
Chapter 9	Does the Authenticity of Preclinical Teaching Format Affect Subsequent Clinical Clerkship Outcomes: a Prospective Randomized Crossover Trial Under editorial review	127
Chapter 10	Using Functional Neuroimaging Combined with a Think-aloud Protocol to Explore Clinical Reasoning Expertise in Internal Medicine Under editorial review	139
Chapter 11	Discussion	159
Chapter 12	Summary Nederlandse samevatting (Summary in Dutch)	179 185
Chapter 13	Appendices	191
Dedications		201
Curriculum vitae		203

Chapter 1

Introduction

Introduction

Expertise in clinical reasoning is important for high quality patient care. Clinical reasoning involves both establishing the diagnosis (diagnostic reasoning) as well as deciding on a plan for action that is specific to the patient's circumstances and preferences (therapeutic reasoning). Indeed, clinical reasoning is part of nearly all aspects of what a physician does in practice.¹ It enables physicians to take 'wise' action, meaning taking the best-judged action in a specific situation or *context*.^{2,3} The complex nature of clinical reasoning has been previously described;¹ it involves, for example, the need to develop a tolerance for ambiguity and a reflective understanding of the practice.⁴ Despite decades of research in clinical reasoning, limited guidance can be given to medical educators regarding how to help learners gain expertise in this situation dependent complex process. Measuring clinical reasoning can be even more challenging.

What is known about the reliable and valid measurement of clinical reasoning in clinical practice?

Several lines of research have focused on the measurement of clinical reasoning. In the 1960s, patient management problems (PMP) were developed to evaluate clinical reasoning. In a PMP, a realistic patient problem is presented to an examinee, with the examinee completing free-text queries. PMP studies demonstrated that a score derived from one problem did not predict performance on another (establishing the concept of the now termed *context specificity*), and experienced clinicians performed similarly to less experienced clinicians or novices.^{5,6,7} This "intermediate effect" may be due, in part, to the inability of PMPs to reflect the *context* of actual clinical practice in which experience makes a difference. These findings cast doubt on the use of this type of assessment tool for measuring clinical reasoning.

Subsequently, it was argued that any clinical problem has one or more features crucial to its diagnosis and management, the so-called "key features approach".⁸ However, results with this approach indicated that performance for clinicians at the end of their training was superior to that of experienced clinicians,⁹ which raised concern about the validity of this approach. Knowledge organization in memory then became a focus for study. Script theory^{10,11} proposes that knowledge organization differs between more and less experienced clinicians; and that such organization enables experienced physicians to efficiently accomplish tasks such as diagnosis and management. In other words, successful diagnosis and treatment involves the organization of knowledge in memory in a way that it can be recalled efficiently and effectively in practice. Work exploring knowledge organization¹²⁻¹⁴ has suggested that physicians actually use multiple cognitive models in practice.¹⁵ These cognitive models are often grouped into analytic (actively collecting and analyzing data to make decisions) and non-analytic (or automatic and based on prior experiences). Research suggests that experts use both cognitive models in different situations,^{16,17} or contexts. However, these cognitive models (analytic and non-analytic) do not account for the impact of context (or the specific presentation of a specific clinical problem in a specific setting) and several investigators^{18,19} have called for the exploration of the impact of context, and the factors that comprise context, on clinical reasoning.

Challenges to the reliable and valid measurement of clinical reasoning: the problem of context specificity

There are many challenges to constructing tools for measuring clinical reasoning in the clinical setting. For example, studies have found that it is difficult to disentangle knowledge, experience, and reasoning.^{1,20} Further, multiple studies have found that performance across situations (cases) can vary greatly--this is often referred to as the problem of context specificity.^{21,22} Indeed, it has been argued that clinical reasoning is revealed only in action;²³⁻²⁴ such action implicitly would incorporate the potential influence of the unique context of the clinical encounter.

Context specificity poses challenges to the measurement of clinical reasoning in several ways. First and foremost, authentic clinical situations (i.e. those involving real patients) can be highly complex and dependent upon the specific patient's problems and setting. Work outside of medicine has found that reasoning in "real-world" situations places significant reliance on the environment.^{25,26} In other words, this work would suggest that the environment plays a critical part in the reasoning process. Thus the clinical *context* may, in part, account for shortcomings in developing reliable and valid tools for reasoning success, both in terms of instructional outcomes as well as expert performance.

Another challenge to addressing the issue of context specificity involves problems with precision of measurement. In essence, we are trying to measure the processes that occur largely inside the physician's mind using tools that capture observable behavior only. As stated above, clinical reasoning is complex in nature. However, there is potential opportunity for improved measurement precision by triangulating traditional behavioral measurements with concurrent verbal reports (think-aloud protocols) or by using novel imaging technology that capitalizes on functional neuroimaging (functional Magnetic Resonance Imaging [fMRI]). Another important reason that the problem of context specificity remains vexed could be that medical educators have not identified and built upon a theoretical framework that captures the multidimensional nature of clinical reasoning in a specific context thus enabling a sequence of studies to be performed to explore this phenomenon. Theoretical frameworks can provide a unique lens through which to view the problem in a new way and can challenge prevailing assumptions that may be erroneous.

The problem of context and measuring clinical reasoning

The influence of context on clinical reasoning is an area that has not been rigorously explored in prior studies. Our understanding has not gone far beyond identifying the problem of context specificity. What is lacking are systematic studies to capture or codify this phenomenon using an inclusive theoretical framework to inform the findings. This gap may be because our current view of context specificity implies "stability" in the individual as well as "stability" in the environment. In this approach the individual is considered to be the unit of analysis (the physician is the source of true variance and reasoning is assumed to reside solely within the decision maker's head), and the additional factors such as the environment and the other participants in the encounter (as well as their interactions) are not a part of the framework as they are considered to be "stable" or "static" or even inconsequential. In other words, the importance of the contribution of the environment, participants (other

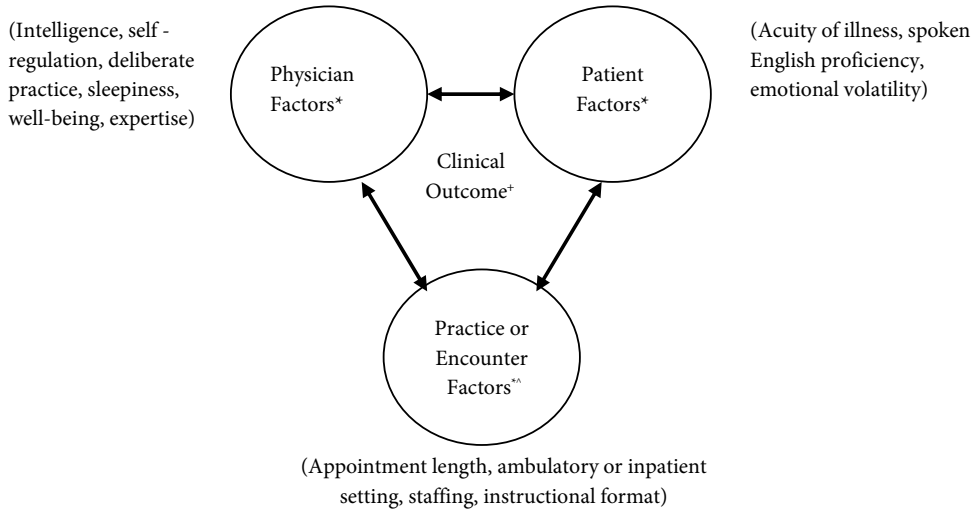
than the decision maker), and their interactions have been largely ignored in medical education. Indeed, potential variance in these areas has been considered to be “noise”, and studies have attempted to control for these factors that may be important components of reasoning “*in vivo*”. These assumptions of stability and the primary unit of analysis should be tested. It is possible that both the problem frame and the solutions are not “static” or “stable”. Furthermore, by considering the physician as the sole unit of analysis, one can fail to capture dynamic interactions between participants in the encounter and the setting.

This thesis begins to explore the influence, if any, of contextual factors on clinical reasoning performance. More specifically, this thesis will address, through a series of studies, the influence of context on clinical reasoning in order to help us better understand, or unravel, context specificity. We will seek to capture the influence of context on clinical reasoning. This thesis will not attempt to measure reasoning from a standard psychometric sense, rather it will attempt to capture or codify clinical reasoning in more naturalistic (“*in vivo*”) settings. This thesis is meant to provide an initial step to addressing the challenge of context specificity and we will do so through the lens of two theoretical frameworks that are novel to many medical educators—situated cognition and ecological psychology. These two theories are useful for this problem as they challenge the assumptions of stability in the physician and the environment as well as the often assumed non-dynamic interactions of the physician with other participants and the environment.

Understanding the influence of contextual factors could ultimately lead to important improvements in patient care as well as more effective teaching practices for helping trainees acquire expertise in this critical area.

A framework for exploring the problem of context in clinical reasoning

As a first step, context was defined as it pertains to the clinical encounter. This definition not only involved clarifying terminology and constructing a view of context for study, it also aimed to propose a theoretical framework that would embrace this view and provide a “roadmap” or “compass” for the systematic study of the influence of context on clinical reasoning. Two chapters of the thesis (2 and 3) will address the central role of theory and, as such, shape the remainder of the investigations. In chapter 2, context is defined as the weaving together (i.e. interaction) of the participants and the environment or setting in the encounter, which creates meaning that importantly evolves. It is the “inter-related conditions in which something exists or occurs”²⁷ (which infers a location or setting) as well as “the parts of a discourse that surround a word or passage and can throw light on its meaning”.²⁷ Thus, context is more than the individual participants in the clinical encounter or the setting where the participants meet to render care. Context emerges from participant and setting interactions (Figure 1) and these cannot be reliably predicted at the outset. Figure 1 illustrates three general factors (physician, patient and practice) as well as specific examples (elements) of each factor. Throughout this thesis, the terms practice and encounter factors are used interchangeably; both terms refer to elements pertaining to the environment or setting for the encounter.

Figure 1: Context in the Clinical Encounter: A Situated Cognition and Ecological Psychology Perspective

*=Individual participants are influenced by affordances, effectivities and goal structures/hierarchy (ecological psychology, discussed below). Examples of each factor (or factor elements) are listed in parentheses next to the factor.

^{^a}=The terms practice factors and encounter factors are used interchangeably throughout this dissertation; both refer to elements pertaining to the environment or setting for the encounter.

⁺=The clinical outcome is dependent on these three factors, their interactions, and possibly other inputs.

Context, then, is the weaving together of participants, within a setting, to create meaning (which evolves). Thus, the outcome of the encounter is based on multiple and interacting inputs, namely participants and setting (Figure 1). In chapter 3, a specific physician factor (or a factor element, see Figure 1), advanced age, is explored and potential implications are proposed. Several issues related to cognitive performance and potential implications associated with aging are outlined. Further, the potential impact of physician aging on the three-factor (patient, physician, and practice) model of context (Figure 1), to include expertise, neurobiology, intelligence, and self-regulated learning, is discussed. Neurobiology and clinical reasoning expertise is also explicitly explored further in chapter 10.

Having defined context in the clinical encounter, we undertook a search for theoretical frameworks that would embrace this definition of context and serve as a foundation for approaching subsequent research questions outlined below. We found two overarching theoretical frameworks which embrace our notion of context in the clinical encounter—situated cognition and ecological psychology.^{26,28-32} These theories argue that knowledge, thinking (reasoning), and learning are situated, or *located*, in experience. Experience refers to not only the participants but also the physical environment or setting, and their interactions. These theories further argue that outcomes in such a system of interactions can be non-linear (the outcome is more complex than the simple sum of the components) as they emerge through these interactions, and are not simply predicted by the presence of the participants taken individually. Notably, in this model, not only the participants but also the

environment (physical setting) is essential; these models concur with our definition of context outlined above.

More specifically, situated cognition embraces the notion of complex interactions between the individual participants and the outcome (in this case patient care) in the medical encounter. Dividing components of the clinical encounter into physician, patient, and practice factors (and the individual elements that comprise each of these general factors), as supported by situated cognition (Figure 1), is critical. Doing so helps to illuminate the potential complex interactions between these factors and how the outcome of the clinical encounter evolves, or can change, based on feedback between each of these factors. Throughout this thesis, we will use the term *contextual factor* to refer to the three components of the three-factor model (i.e., physician, patient, and practice). In addition, we will refer to the individual elements within each of these factors as specific or selected contextual factors (see parentheses in Figure 1). For example, *well-being*, *age*, and *self-regulation* would all be considered specific or selected contextual factors within the more general physician factor.

Adding to the situated cognition view, ecological psychology proposes the important dyad of the agent-environment interaction in terms of affordances (what the environment provides) and effectivities (what the participant can potentially do in that environment). These two concepts (affordances and effectivities) cannot be completely separated (they are interdependent) and are predicated on the specific situation. A car is drivable for adults because they have the effectivity, but a car is not drivable to an infant. Further, ecological psychology, with its focus on agent-environment interactions, argues that outcomes emerge from *goal-driven* participants interacting with each other in an information-rich environment. Participants can hold multiple goals that can change as the encounter emerges. These changing goals are proposed to act as a driving force for how individual participants act in a given context.

Thus, the situated cognition and ecological psychology perspectives illustrated in Figure 1 propose that the environment (represented by the encounter or practice factor) is a part of the signal, or meaningful variance to be measured, as opposed to “noise” to be ignored. In other words, from these viewpoints, one cannot fully appreciate reasoning without considering the *experience*, since reasoning is located in experience (participants, environment, and their interactions). This notion, provides a framework to explore or unravel context specificity.

Research questions addressed in this thesis

The overarching research question in this thesis is whether selected contextual factors impact clinical reasoning performance, both in expert practice and during the instruction of trainees.

Specific research questions were:

1. What is the influence of contextual factors in expert practice?
2. Does the authenticity of instructional format impact expertise development in clinical reasoning?

3. Can we improve upon assessment of clinical reasoning by triangulating behavioral measurements with concurrent verbal reports (think-aloud protocols) or with novel instruments such as fMRI scanning?

Research Question 1: What is the influence of contextual factors in expert practice?

For the first research question, we conducted two thematic reviews, which are detailed in chapters 2 and 3 of this thesis, as we believed that theoretical frameworks were needed to explore the multidimensional and complex construct of context in the clinical encounter. These chapters provide a “roadmap” or “compass” for the subsequent investigation of the influence of context on learning and clinical practice.

The implications from these reviews are twofold:

1. The outcomes of clinical encounters can be complex and potentially nonlinear, and this has implications for data analysis in subsequent studies. In other words, the results of interactions of factors in Figure 1 may be more than additive. This may help explain why so many predictive validity studies in medical education fail to capture much variance in the outcome measure of interest (see references in chapter 2 of this thesis) and it also suggests that we may need to consider other mathematical models when exploring clinical reasoning.
2. Studying specific, “extreme” forms of individual patient, physician, and practice factors (and their interactions) to unravel the influence of context on diagnostic and therapeutic reasoning. By studying extreme forms, the influence, if any, on reasoning could potentially be observed with a smaller sample size.

Next, the perspectives of medical students, patients, and physicians viewing the same event were explored to see if three different stakeholder groups view the same event differently, which would add validity evidence to our proposed theoretical model. This study (chapter 4) involved a post-encounter form (chapters 4 and 5) that was also designed to be used in future investigations. If these theoretical constructs have applicability to the clinical encounter, one would anticipate that these three stakeholder views of the same event would have quite different viewpoints, as well as some similarities, as all three stakeholder groups have different affordances and effectivities (chapter 4). Or stated another way, if context influences reasoning through factors shown in Figure 1, one would expect that different stakeholder groups (physician, student, and patient) could see the same event differently because of different views regarding these factors. Establishing the feasibility, reliability, and validity of this post-encounter form was also investigated (chapter 5), given its use in future studies. Ultimately, a version of this form (chapter 5) was used in subsequent experiments detailed in research questions 2 and 3.

As these investigations provided support for our proposed theoretical model, videotapes were then constructed to conduct a series of experiments employing various “extreme” patient, physician, and practice factors to study if these specific factors influence practicing physicians’ clinical reasoning (diagnostic and therapeutic). Videotapes were used as they allowed us to consistently reproduce the presence of the specific contextual factor (or not) while holding other independent variables constant. We explored specific physician, patient, and practice factors, separately and in combination, in this investigation. We

selected specific individual factors for study, alone and in combination, as a starting point to addressing how context may influence clinical reasoning in this complex environment. See chapters 6 and 7 for an overview of the effect these specific factors had on physician clinical reasoning strategies and performance.

Research question 2: Does the authenticity of instructional format impact expertise development in clinical reasoning?

In this series of investigations, the relationship between authenticity (approximation to “real-life” or *in vivo* experience in actual practice) of instructional format and selected outcome measurements (scores on written and clinical skills exams, as well as performance on the internal medicine clerkship) was explored. This research question focused on the balance between increasing the authenticity (and potentially bolstering emotional engagement that might lead to improved learning) of the instructional format versus negatively impacting cognitive load with increasing authenticity of instruction, which could result in decrements in learning. We wished to study the development of expertise in students and thus conducted a prospective, randomized crossover trial spanning a two-year period. This research question helps address the problem of context specificity from the standpoint of learning clinical reasoning.

The balance between increasing authenticity (a patient factor, as well as a patient and physician *interaction* factor with the latter represented by the bidirectional arrow in Figure 1) and cognitive load (a physician factor) is an important area for investigation according to situated cognition and ecological psychology theories (see Figure 1). These theories would acknowledge that both are important, and that the effect of these opposing views is not known (and thus worthy of exploration). These theories would argue that the balance between increasing authenticity (emotional engagement) versus increasing cognitive load is likely dynamic (due to interactions between participants and their environment). These theoretical views contrast with the individual views of cognitive load and emotional engagement theories, respectively, that would argue that the answer is likely “either/or” in these more linear perspectives. Situated cognition and ecological psychology theories would also argue that the outcome would be dependent upon the setting which was held relatively constant in this study. The results of the preclinical investigation are illustrated in chapter 8 and the results of the clinical investigation are found in chapter 9.

Research question 3: Can we improve upon assessment of clinical reasoning by triangulating behavioral measurements with concurrent verbal reports (think-aloud protocols) or with novel instruments such as fMRI scanning?

This research question involved instrument development work. Specifically, as clinical reasoning is only partly visible in action (because the neurocognitive processes in the physician’s head have not been measured in prior work), we sought to explore this largely “hidden” process through two means: fMRI imaging and think-aloud protocols. The latter technique was used in several investigations (see chapters 6, 7, 10) and the former technique was explored to see if fMRI imaging could serve as means of understanding clinical reasoning expertise (chapter 10). Specifically, we used a gold standard of physician expertise in the

United States (performance on high-stakes, multiple-choice questions for certification) and compared findings on these items with a neuroimaging correlate (fMRI scanning) while reading, answering, and reflecting on these items (see chapter 10).

Concurrent verbal reports (think-aloud protocols used in this thesis; chapters 6, 7, 10) were also combined (triangulated) with a behavioral measurement in several investigations, namely a post-encounter form (see chapters 4-7). The think-aloud protocol data provides concurrent (while performing the task) verbal information to reflect what the physician is thinking as they decide and act (the final decision and action in our studies was reflected in the post-encounter form). Think-alouds could provide important concurrent information (patient, physician, practice factors and their interactions) to help validate situated cognition and ecological psychology theory in clinical reasoning. In the fMRI study, which explored the functional neuroimaging of clinical reasoning expertise, a think-aloud protocol was also applied (see chapter 10). This was done to add important verbal-report information to standard, accepted, multiple-choice examinations. We began our exploration of functional neuroimaging (chapter 10), which measures relative regional brain perfusion, controlling for both environmental and patient factors (multiple-choice questions versus live patients). We did this because work on fMRI and clinical reasoning had not been performed, and we wished to establish fMRI findings for the physician factor before exploring fMRI findings with altering patient and practice factors. By obtaining a very large number of images over a set period of time, the fMRI perspective also offers a potential alternative to the standard “linear” approaches to measuring the complex nature of clinical reasoning that could inform future work. We also hope that this fMRI work could expand to more “*in vivo*” situations, where physician, patient, and practice factors likely interact more dynamically, thus impacting, to an even greater extent, clinical reasoning (chapter 10).

If context impacts clinical reasoning, and the theoretical frameworks that were chosen are demonstrated to apply in medical education settings, an important implication is establishing the need to better understand complexity and complexity theory. Our theoretical frameworks would suggest that, in some cases, the interactions could lead to nonlinearity and potentially even deterministic chaos (chapter 2). If clinical reasoning in certain practice situations can be shown to demonstrate nonlinearity (chapter 2) then our current statistical analysis techniques are not sufficient for capturing clinical reasoning in certain situations. Determining when nonlinearity may occur would be a fruitful endeavor and has been shown to be difficult to predict in other fields such as meteorology.³³ We will need to learn to embrace nonlinear approaches to questions, that could at times demonstrate nonlinearity, like the influence of context on clinical reasoning, to better understand these essential phenomena for rendering high-quality patient care.

Disclaimer: As the thesis chapters are written as stand-alone manuscripts for the purpose of peer-reviewed publication, there is necessarily some redundancy in the text of this dissertation.

References

1. Higgs J, Jones MA, Loftus S, Christensen N. *Clinical Reasoning in the Health Professions*. 3rd ed. Oxford: Butterworth Heinemann; 2008.
2. Cervero RM. *Effective Continuing Education for Professionals*. San Francisco: Jossey-Bass; 1988.
3. Harris IB. New expectations for professional competence. In: Curry L, Wergin JF and Associates, eds. *Educating Professionals: Responding to New Expectations for Competence and Accountability*. San Francisco: Jossey-Bass; 1993:17-52.
4. Bleakley A, Farrow R, Gould D, et al. Making sense of clinical reasoning: judgment and the evidence of the senses. *Med Educ*. 2003;37:544-552.
5. Schmidt HG, Boshuizen HPA, Hobus PPM. Transitory stages in the development of medical expertise: the 'intermediate effect' in clinical case representation studies. In: *Proceedings of the 10th Annual Conference of the Cognitive Science Society*. New Jersey: Erlbaum; 1988.
6. van der Vleuten CPM, Luijk SJ, van Beckers HJM. A written test as an alternative to performance testing. *Med Educ*. 1989;23:97-107.
7. van der Vleuten CPM, Newble D, Case S, Holsgrove G, McCann B, MacRae H, et al. Methods of assessment in certification. In: Newble D, Jolly B, and Wakeford R, eds. *The Certification and Recertification of Doctors: Issues in the Assessment of Clinical Competence*. Cambridge (MA): Cambridge University Press; 1994.
8. Bordage G, Page G. An alternative approach to PMPs: the 'key features' concept. In: IR Hart, RM Harden, eds. *Further Developments in Assessing Clinical Competence*. Montreal: Heal- Publications; 1987:59-75.
9. Bordage G, Brailovsky CA, Cohen T, Page G. Maintaining and enhancing key decision-making skills from graduation into practice: an exploratory study. In: AJJA Scherpbier, CPM van der Vleuten, JJ Rethans, eds. *Advances in Medical Education*. Dordrecht, The Netherlands: Kluwer Academic; 1996:128-130.
10. Feltovich PJ, Barrows HS. Issues of generality in medical problem solving. In: Schmidt HG and De Volder ML, eds. *Tutorials in Problem-based Learning: A New Direction in Teaching the Health Professions*. Assen: Van Gorcum; 1984:128-142.
11. Schank RC and Abelson R. *Scripts, Plans, Goals, and Understanding*. New Jersey: Erlbaum; 1977.
12. Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise: theory implications. *Acad Med*. 1990;65:611-621.
13. Barrows HS, Feltovich PJ. The clinical reasoning process. *Med Educ*. 1987;21:86-91.
14. Custers EJ, Regehr G, Norman GR. Mental representations of medical diagnostic knowledge: a review. *Acad Med*. 1996;71:S55-S61.
15. Norman G. Research in clinical reasoning: past history and current trends. *Med Educ*. 2005;39:418-427.
16. Elstein AS, Shulman LS and Sprafka SA. *Medical Problem-solving: An Analysis of Clinical Reasoning*. Cambridge (MA): Harvard University Press; 1978.
17. Eva K. What every teacher needs to know about clinical reasoning. *Med Educ*. 2005;39:98-106.
18. Gruppen LD, Frohna AZ. Clinical reasoning. In: Norman GR, Newble DI, van der Vleuten CPM, eds. *International Handbook of Research in Medical Education*. Dordrecht: Kluwer; 2002.
19. Norman GR, Tugwell P, Feighter JW, Muzzin LJ, Jacoby LL. Knowledge and clinical problem solving. *Med Educ*. 1995;19:344-356.
20. Ericsson KA, Charness N, Feltovich P, Hoffman RR, eds. *The Cambridge Handbook of Expertise and Expert Performance*. New York: Cambridge University Press; 2006.
21. Eva K. On the generality of specificity. *Med Educ*. 2003;37:587-588.
22. Eva KW, Neville AJ, Norman GR. Exploring the etiology of content specificity: factors influencing analogic transfer and problem solving. *Acad Med*. 1998;73:S1-S5.
23. Charlin B, Tardif J, Boshuizen HPA. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad Med*. 2000;75:182-190.
24. Brailovsky C, Charlin B, Beausoleil S, Coté S, van der Vleuten C. Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency: an experimental study on the script concordance test. *Med Educ*. 2001;35:430-436.
25. Chaiklin S and Lave J. *Understanding Practice: Perspectives on Activity and Context*. New York: Cambridge University Press; 1993.
26. Lave L and Wenger E. *Situated Learning: Legitimate Peripheral Participation*. New York: Cambridge University Press; 1991.
27. Webster's online dictionary. Context. Accessed 12/1/2010.

28. Wilson BG, Myers KM. Situated cognition in theoretical and practical context. In: Jonassen D and Land S, eds. *Theoretical Foundations of Learning Environments*. New Jersey: Erlbaum; 1999.
29. Heft H. *Ecological Psychology in Context: James Gibson, Roger Barker, and the Legacy of William James' Radical Empiricism*. New Jersey: Lawrence Erlbaum Associates; 2001.
30. Young MF. Instructional design for situated learning. *Educ Technol Res Dev*. 1993;41:43–58.
31. Bredo E. Reconstructing educational psychology: Situated cognition and Deweyian pragmatism. *Educ Psychol*. 1994;29:23–35.
32. Robbins P and Aydede M. *The Cambridge Handbook of Situated Cognition*. New York: Cambridge University Press; 2009.
33. Jonassen DH and Land SM. *Theoretical Foundations of Learning Environments*. New Jersey: Lawrence Erlbaum Associates; 2000.

Chapter 2

Redefining *Context* in the Clinical Encounter: Implications for Research and Training in Medical Education

Published in:

Academic Medicine 2010;85:894–901

Steven J. Durning, Anthony R. Artino Jr., Louis N. Pangaro, Cees van der Vleuten, and
Lambert Schuwirth

Abstract

Physician training and practice occur in complex environments. These complex environments, or contexts, raise important challenges and opportunities for research and training in medical education. The authors explore how studies from fields outside medicine can assist medical educators with their approach to the notion of context in the clinical encounter. First, they discuss the use of the term context in the clinical encounter as it relates to medical education. They then detail the meaning and use of the term in diverse fields outside medicine, such as mathematics, physics, and psychology, all of which suggest a nonlinear approach to the notion of context. Next, the authors highlight two inclusive theories, situated cognition and ecological psychology, that propose factors that relate to context and that suggest some potential next steps for research and practice. By redefining context as it relates to the clinical encounter (by linking it to theory and research from several diverse fields), the authors hope to move the field forward by providing guidance for the theory, research, and practice of medical education.

Introduction

Physician training and practice occur in many complex environments ranging from classrooms to busy inpatient wards, emergency rooms, operating rooms, and ambulatory clinics. These environments, or contexts, influence how physicians learn, teach, and practice medicine and, thus, have important implications for research and training in medical education. In this article, we explore the use of the term context in the clinical encounter as it relates to medical education. We begin with examples of challenges related to the current view of context in the clinical encounter, and then we redefine context and detail the meaning and use of the term in diverse fields such as mathematics, physics, and psychology. The theories we discuss illustrate how individual system components, or factors, can relate to the overall system in complex ways. Next, we highlight several of these factors and their interactions in the clinical arena. We contend that it is essential to understand how such factors may relate in the complex clinical environment, resulting in good or possibly suboptimal patient care. Further, we outline two additional inclusive views of context (situated cognition and ecological psychology) that highlight challenges and implications in the field of medical education. A key point is that these theories argue for nonlinear approaches, where the sum (or outcome) can equal more than the individual factors. Finally, we propose next steps for research regarding context as it relates to medical education. Ultimately, our goal in this article is to reframe the definition of context in the clinical encounter by relating it to findings from several other fields, first from an individual participant perspective and then from an overall system perspective. In doing so, we hope to provide some guidance for the theory, research, and practice of medical education.

Challenges Related to Current Views of Context: Outcomes and Terminology

Knowing what contributes to the outcome of a clinical encounter is essential to improving medical education and, ultimately, patient care. Medical educators have investigated several factors (as defined above) in this complex system, but largely in isolation. For example, researchers have explored the impact of one or more factors on system outcomes, including attempts to predict professionalism,¹⁻⁴ clinical skills,⁴⁻⁷ communication skills,^{4,6,8} and even overall competence.^{4,8,9,10} Although such studies can illuminate the importance of individual factors, the limitation of approaching factors in isolation is evident when one explores the association between inputs and outcomes. For instance, numerous predictive validity studies in medical education that have addressed individual system factors, such as professionalism or clinical skills, have found a consistent theme: selected inputs tend to explain little variance in outcome(s).¹⁻¹⁰ Stated another way, medical education research has been limited by not yet exploring interactions between system components, interactions that other fields have demonstrated are complex. Such exploration involves, in part, considering nonlinear approaches that do not assume a fixed, linear relationship between two or more variables. The restricted findings from the field of medical education are not unlike those from the human intelligence literature, which, for decades, has attempted to explain the rise in intelligence (IQ) testing with individual system factors like parental IQ and type of school, but with little success.^{11,12} If acquiring and demonstrating competence in medicine is similar to

acquiring and demonstrating intelligence, then recent findings from the intelligence literature, which argue for nonlinear (or multiplier) approaches, could help inform the theory and practice of medical education.¹¹

An additional challenge to medical education research is the use of multiple terms that essentially mean the same thing, or, alternatively, give multiple, potentially conflicting meanings to the same term. The use (or misuse) of these multiple terms interferes with referencing prior medical education literature and/or literature from other fields. One example of this phenomenon is the use of the term context in medical education.

Current Views on Context

It is useful to begin with common meanings of this term. In the medical education literature, there are several definitions, which fall into two general categories. Context may refer to the setting or physical location of the encounter (such as ambulatory, ward, emergency department, and operating room). In this sense, context answers the question “Where does the problem take place”? Although different locations of practice do reflect context, we do not believe that such a definition is sufficiently inclusive to inform medical education and research.

Context may also refer to the content of a clinical problem, such as additional information provided in a test question that is related to a presenting problem. From this perspective, a “context-rich” multiple-choice question includes many details of a clinical presentation, whereas a “context-poor” question does not include (or includes very little) patient information. Likewise, routine information collected after the history of present illness is typically thought of as context. Although such additional details regarding the patient (or problem) do provide information about context, this definition, also, is not sufficiently inclusive.

Redefining Context

The word context is derived from the Latin *contextere*, meaning “to weave together”. Webster’s online dictionary defines context as the “interrelated conditions in which something exists or occurs” with environment as a synonym.¹³ This definition can be equated with location or setting. A second definition of context is “the parts of a discourse that surround a word or passage and can throw light on its meaning”.¹³ Because in medicine this discourse includes information that accumulates during the time of a patient encounter, the term context is not restricted to a static situation, but instead is one that evolves. As such, context (1) comprises interacting factors that add to the meaning of something that exists or occurs in an environment and (2) allows for change in that meaning as information is added over time. The factors that “weave together” the medical encounter involve the participants and their interactions. The participants are the patients and physicians (and, at times, additional individuals such as family members of the patient, medical students, nurses, and/or resident physicians); the interactions include how participants relate to each other as well as to the physical location of the visit. Interactions are likely shaped by factors unique to the individual participants, such as expectations, goals, communication skills, and well-being, as well as participants’ shared communication or reactions to each other. The end result, or “suc-

cess,” of the encounter depends on these unique participant factors and their interplay—in other words, the outcome of the clinical encounter likely has multi-factorial inputs that interact. Thus, there is a view of context in the clinical encounter from the individual participant and another view from the entire system (i.e., how all the components in the system interact to affect the individual participant and overall outcome, respectively). In this article, we address both views and propose a comprehensive model for approaching research that goes beyond detailing items in the clinical context and emphasizes the importance of their interaction. We explore both views through the theoretical lenses of fields outside medicine, asking how these fields can inform the various system components, or factors — patient, physician, and setting—that constitute the clinical context.

Context involves not only location and interactions between participants but also time. Context, considered as the interaction of clinical information, changes over time as clinical information accumulates—for instance, the clinical encounter develops or progresses in time. Considering context as changing over the time of the encounter further illuminates the idea that the outcome of a clinical encounter is not static and its result is not merely dependent on the patient or the physician, for that matter. Instead, the outcome of a clinical encounter is dependent on those factors (patient and physician) and on how the patient and physician interact with respect to a specific problem in a specific location in a specific unit of time. More specifically, on reflection, it is clear that the outcome is dependent on a variety of system factors: factors unique to the patient, factors unique to the physician, factors unique to the physical location or setting of the encounter and, finally, factors that represent their interactions (patient-physician, patient-setting, physician-setting, and patient-physician-setting). Further, these factors are likely to interact in complex ways. In fact, the notion that these factors interact in complex ways implies that, and perhaps helps one understand why, algorithms, checklists, and guidelines may only have limited efficacy in health care. The interplay between patient, physician, and practice setting is sufficiently complex and unique as to limit the effective use of such relatively stringent and fixed criteria in all but the simplest cases.

In the next section, we discuss theories from other fields that provide insight into the complex system of medicine. These theories emphasize nonlinearity, or the idea that the sum (or outcome) may equal more than the individual parts. We begin with mathematics and quantum physics. Then, we discuss two inclusive theories (situated cognition and ecological psychology) that incorporate the different contextual factors in medical education; these theories approach context from the view of the entire system. Next, we discuss theories that illuminate how context may influence the three individual factors (patient, physician, and encounter) as proposed by situated cognition and ecological psychology. Finally, in the last section, we propose some potential implications for research.

Insights from Other Fields

Context and Mathematics (Chaos Theory)

Chaos theory is a prototypical, nonlinear, mathematical model. In much of science, processes have been described from the view of linear models, which implies that if we have

enough measurement points we can predict precisely the situation of an object at any moment. Nonlinear models, like chaos theory, argue that because of interactions between components in a system, the results can accumulate in nonlinear (other than additive) ways, making precise prediction of the outcome impossible at the onset. Indeed, it is likely that most social processes are subject to a constant interaction of (opposing) forces that render such a precise determination impossible, no matter how many data points we collect. Chaos theory should be differentiated from the term “chaos”, which is often used in the popular press to mean things that are completely unpredictable and purely subject to chance. Chaos theory does not argue for randomness; instead, this mathematical theory states that outcomes are bounded, meaning that although an outcome cannot be precisely predicted at the onset, there are limits to the possible interactions.

Chaos, as described in chaos theory, has been found to apply to several complex systems.¹⁴⁻¹⁶ Indeed, some have argued that nearly all social interactions, such as patient care, are chaotic¹⁶ because social interactions evolve with participant feedback in unique and dynamic ways. A number of terms that are associated with chaos or are synonymous with chaos exist in the literature, depending upon the field applying the theory, further contributing to confusion. These terms include complex complexity, dissipative systems, dynamical systems, catastrophe, bifurcation, and emergence.¹⁶

A key principle of chaos theory is that any outcome has “sensitive dependence upon the initial conditions.” That is, factors can interact in a nonlinear fashion, which then has an impact on the results, or outcome, of an encounter. For example, chaos theory would argue that a medical student who has exceptional reading skill (the “outcome”) likely had several factors (the “initial conditions”) that led to this result: innate reading skill, encouraging parents, and interest and motivation to read (with positive reinforcement between each of these other factors). The encouragement of the parents may have made the student more open to role modeling by a well-read physician and this, in turn, may have influenced the receptiveness of the student to encouragement by the parents (a powerful reciprocal interaction, or feedback loop). The resulting “exceptional reading skill”, then, is likely more than the sum of the individual inputs, or factors; instead, the result emerges also from nonlinear interactions.

From the chaos theory approach, the determination of the outcome is more than simply including more measurement points in the model. Although including more measurements can improve understanding, from the lens of chaos theory, we need to also explore complex interactions between such measurements. Chaos theory highlights the perceived importance of certain parameters, or factors, initially perceived as irrelevant that become relevant when we explore the unique contribution of each factor and their interactions in a nonlinear fashion. This is not to say that exploring linear relationships is not important. Moreover, addressing “extreme phenotypes” or “classic examples” of system components can facilitate our understanding of not only the relationship of the component to the outcome but also our understanding of how the measure may dynamically interact with other components in the system. However, if we want to predict, in an individual case, how a specific factor (e.g., patient satisfaction) interacts with other factors in the system (e.g., physician well-being) at a given moment of space and time, this literature would suggest that nonlinear approaches are optimal. As another example, consider slowly pouring a bottle of sand onto the floor. The sand will form a pile, but with closer inspection we will

see “avalanches” of sand grains from time to time. Predicting the path of a single grain of sand in such an avalanche is not possible because it represents a chaotic system (it involves many, many interaction effects as grains of sand bump into each other). The shape of the sand pile is dependent upon the initial conditions such as the size of the bottle of sand, speed of pouring, and wind; however, prospectively predicting where each grain of sand will lie in the pile is not possible. The sand pile also raises the difference, mentioned earlier, between chaos theory and the term chaos as typically used—chaos theory argues that the results are bounded by the initial conditions, and although the options are infinite and cannot be precisely predicted, conditions can be roughly predicted and can also be understood with retrospective analysis.

In terms of a medical encounter, a potential scenario could be a tired and overworked resident seeing a frustrated patient in a busy outpatient practice. The fact that the patient has had to wait to be seen, is upset by the lack of clarity in prior diagnosis, and is frustrated by interacting with a resident who needs to repetitively restate the patient’s concerns due to his own fatigue (and who becomes easily angered due partly because of fatigue and time pressure in clinic), likely leads to nonlinear results in the patient’s lack of satisfaction with the visit (an undesirable outcome). These factors “bump into each other” through many different interaction effects, much like the grains of sand in the example above. More specifically, the successful physician cannot merely approach the clinical encounter with a relatively inflexible guideline or algorithm that he or she applies to all cases and expect success in the complex practice environment. That would be like expecting the sand pile to be identical each time sand is poured from the bottle. Instead, he or she needs to be aware of factors (patient, physician, setting, patient and physician, patient and setting, physician and setting) that will interact in complex and frequently unexpected ways. In such a system, the physician must be flexible, have some responsiveness to these factors, and be willing and able to adapt to the factors and their interactions.

Context and Quantum Physics

Quantum physics provides an additional lens for viewing context and the interaction of elements in a clinical encounter. In quantum physics, the Heisenberg uncertainty principle¹⁷ states that certain dynamic physical quantities, like position and momentum, cannot both have precise values at the same time. To illustrate this point, consider the movement of a pendulum. Standard physics is able to describe only where the pendulum begins and where it ends; knowing (or predicting) where the pendulum will be at a given moment in time and space is not possible.

This principle has important implications for medical education and medical education research. The argument here is that it is not possible to predict the outcome in a system with multiple changing parts, such as a clinical encounter. Regardless of the number of measurements, the outcome will still be unpredictable and can be prospectively determined only roughly in space and time. In short, system complexity affects our ability to account for variance in the outcomes.

We propose the movement of a pendulum is analogous to the clinical encounter. We know the starting point (the beginning of the visit) and the ending point (desirable and undesirable clinical outcomes), but what happens in between is uncertain at a given mo-

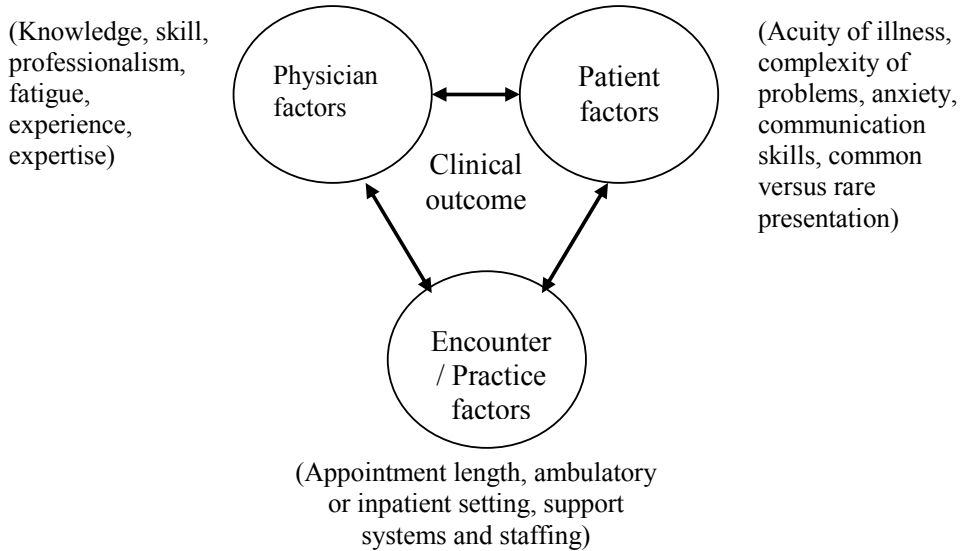
ment of time (it depends upon multiple, nonlinear interactions between the factors described). Or, if we return to our example of the grains of sand, these factors are bumping into each other, with too many possible interactions and feedback between these interactions, for the outcome of those factors to be predicted with exacting precision at the outset. In the following subsection, we approach context in medicine from the perspective of two theories that provide a more unified approach to the notion of context in medicine--situated cognition and ecological psychology.

Context and Situated Cognition

Situated cognition theory contends that all thinking and learning must be viewed as “situated” within the larger physical and social context of the environment. Situated cognition views “cognition, knowing, and learning as all taking place as interactions between people and their environment”.¹⁸ Thus, situated cognition embraces the notion of a complex interplay between participants and their environment as leading to desirable (or undesirable) results. Situated cognition theorists, however, go even one step further, putting equal emphasis on the environment and the person.¹⁹ Indeed, the stand-out characteristic of situated cognition is the placement of individual cognition within the larger physical and social context of interactions and culturally constructed tools and meanings.¹⁸ Situated cognition shifts focus from “individual in environment to individual and environment”.²⁰ What is more, situated cognition sees “writing, conversing, or thinking as the result of dialogue in which the person and environment (ideally) modify each other so as to create an integral performance. It is the production of a well-coordinated performance that involves a kind of dance between person and environment rather than the one-way action of one on the other”.²⁰

Viewing a medical encounter from a situated cognition perspective is enlightening. For example, even if a physician is familiar with a case ahead of time, he or she does not enter that patient encounter knowing exactly how the encounter will proceed. Instead, the “success” of the encounter is determined not by the doctor, not by the patient, and not by the physical environment, but by the dynamic interactions that occur between the doctor, patient, and environment (interactions that are far from invariant across doctor, patient, and environment). Consequently, we embrace the situated cognition viewpoint as a way of reframing the notion of context in the clinical encounter, because such a viewpoint does more than just acknowledge the environment, it places the environment on equal footing with the people in that environment and their unique interactions. Figure 1 illustrates a situated cognition approach to the medical encounter. However, although situated cognition theory does provide a model for viewing the component parts of the medical encounter and their interrelations, it does not illuminate the question of participant goals and available resources in the complex system. We believe that this important additional information can be viewed through the lens of ecological psychology (discussed below), which focuses on further understanding the interplay of humans and their surroundings.

Figure 1: A situated cognition approach to the clinical encounter. Situated cognition embraces the notion of complex interactions between the individual participants and the environment, all of which can influence the outcome (patient care) in the medical encounter model shown in the figure. The three factors that make up the encounter are all-important and interact, as shown by arrows. Encounter and practice factors are used interchangeably and refer to elements pertaining to the location or setting of the appointment. Examples of elements of each factor are listed in parentheses next to the factor. The clinical outcome is dependent upon these three factors, their interactions, and possibly other inputs. Because the clinical encounter evolves (or can change on the basis of feedback between these elements), the outcome(s) can be nonlinear (deterministic chaos).



Context and Ecological Psychology

Ecological psychology prefers an integrated, agent-environment view of participation. This theory embraces the use of mathematical models emerging from chaos theory for understanding complex systems and how such models may be applied to other fields, such as medical education. Ecological psychology stresses the impact of the goals of participants as the driving behavior in complex systems. The environment gives functional value, or “opportunities,” to participants. These opportunities, referred to as affordances, are what the environment offers to participants, what it provides or furnishes, either for good or ill.²¹ For example, a doorknob provides the affordance of “turnable” and a lake the affordance of “swimmable”. Affordances are possibilities for action, and they are not stable for all participants at all times. Affordances exist only for certain classes of participants and are important only when a potential need (or intention) arises. A participant’s skill and ability to act in ecological psychology are termed effectivities.

Thus, ecological psychology informs us about context through the notions of affordances (possibilities), effectivities (ability to use affordances), and intention (or desire to glean an affordance, which is directed by a participant’s goals). From the ecological psychology perspective, obtaining and maintaining competence is accomplished through the education of intention and attention; the former refers to previously constructed or emerg-

ing goals of the encounter and the latter to the process of tuning one's perception to detect potential environmental affordances.

Like theories of situated cognition, ecological psychology provides additional perspective on the dynamic factors and relationships that influence context. Furthermore, ecological psychology implores educators to be aware of these factors and their interrelations when teaching, learning, and conducting research. For example, suppose a patient presents to the Emergency Department with shortness of breath and absent lung sounds on the left side, then rapidly goes into a pulseless electrical rhythm. The technician pages the physician because he or she lacks the effectivity to arrive at the diagnosis. The physician rapidly diagnoses and contemplates the immediate treatment of the tension pneumothorax—a goal that emerges and was not evident on presentation. She calls for a 12-gauge needle (affordance), but then notices a 14-gauge needle on the crash cart (attention). Because she understands the condition and how to act (effectivity), she rapidly inserts the needle into the patient's left chest wall (intention), resulting in a return to sinus rhythm and normal breath sounds. Another example could be a patient presenting with low back pain that radiates down the back of the leg. The physician considers a straight leg raise as a part of a focused physical exam (affordance), but the patient cannot lie flat (attention). The physician thus raises the patient's leg while sitting in a chair (intention), reproducing the patient's symptoms and confirming the diagnosis (effectivity). In this example, the starting point is known as well as the desired outcome; however, the various aspects of the encounter—affordance, attention, intention, and effectivity—emerge dynamically. These various aspects are due to evolving interactions with the patient, the doctor, and the setting and thus cannot be precisely predicted at the outset. Therefore, the outset is known as well as the desirable outcomes, at least in an approximate sense; however, what happens in between is not known, similar to the situation in the pendulum and the grains of sand examples.

We now discuss context from the viewpoint of theories that address individual system components (or individual factors) of context and that are consistent with situated cognition and ecological psychology theories.

Context and Situation Awareness

Situation awareness is understood as an up-to-date understanding of the world around the observer.²² Formally stated, situation awareness is defined as “the perception of the elements in the environment within a volume of space and time, the comprehension of their meaning, and the projection of their status in the near future”.²² This framework requires a more comprehensive view of context because the framework considers the complex interplay between the environment's elements within a volume of space and time. That complex interplay illuminates how the magnitude of the interaction between factors influences the outcome. This theory considers context from the unit of analysis of the participant (participant's view) and could reflect either the patient or the physician in Figure 1.

Situation awareness can also inform decision-making and performance in the context of a medical encounter. Environmental features that affect how well individuals perform include the capability of the system to provide the needed information (e.g., limited time in busy practice), the design of the system interface that determines which information is available (e.g., availability of records), system complexity (e.g., number of patients, number

of problems, support), the level of automatization present in the system (e.g., support from the front desk and nurses and triage of appointments), stress, and workload. Furthermore, situation awareness includes a number of factors internal to participants, factors that have an impact on performance in space and time. These include perceptual processing and limited attention, limited working memory, expectation bias, and pattern-matching to existing schemas.²³⁻²⁵ Attention to these features, and how they might interact, has far-reaching implications for research and teaching.

Context and Expertise

A growing body of literature suggests that deliberate practice is needed for the development of expertise and expert performance. Deliberate practice is the effortful focus on specific aspects of training and often includes feedback from a coach or mentor for the purpose of improving performance. Many fields suggest that a minimum of 10 years of deliberate practice is needed to develop expertise.²⁶ To become an expert in nonlinear dynamic processes, like the practice of medicine, simple instruction is unlikely to be enough. Individuals also need a wealth of experience through which evaluation of multiple iterations of a potential diagnosis are seen so that probability can be assessed reasonably. Working with a coach or mentor to improve performance is more efficient than learning on one's own. Returning to our sand pile example, if we study 100,000 piles of sand, we can more accurately guess shape of the final pile (though we cannot precisely predict the position of individual grains of sand). Using our pendulum example, if we study 100,000 swings of a pendulum, we can more accurately guess where the pendulum will begin and end, though we cannot precisely predict its location at a specific moment of time.

By parsing performance into components so that effortful practice and improvement can be obtained on specific skills, context is implied. For example, to become an expert chess player, a deliberate practice activity might involve studying the chess moves of other experts in chess. This process involves taking into account numerous chess board positions and potential future positions (a practice of conceptualizing immediate and future contexts). What is not explicit in the current theory of deliberate practice is the idea that the context of medicine is not bounded like the relatively static chess board—there is the additional complexity of the practice setting component. This requires that the physician use various cognitive and psychomotor skills to meet the needs of the patient at the point of care. The medical context also involves the additional interplay of multiple participant factors. Educational models recognize that different areas of expertise may require different approaches to deliberate practice (e.g., a concert pianist will have a different deliberate practice approach than a sprinter or chess master).^{26,27} This more inclusive view recognizes that explicit physician skills may vary based on prevailing system component factors (e.g., patient needs) and thus optimal deliberate practice activities may differ on the basis of what is required in the specific patient care situation.

Another important distinction between medicine and other fields that require expert performance is the idea that, in medicine, no two patients with the same disease are identical. This raises an additional issue of context—how a patient presents in a given setting and how that patient interacts with the physician. The theory of contextual interference argues that to become an expert, a learner needs practice dealing with the problem (e.g., a present-

ing symptom or finding) in different environments (e.g., different practice settings) and with different levels of difficulty (e.g., common versus atypical presentation and acuity of presentation). Thus, context appears to be important for developing expertise, yet we know little about how modifying different system components or factors (e.g., a patient suggesting a diagnosis or therapy) might affect the clinical outcome. Additionally, there is the issue of arrested development (i.e., not reaching the level of expert) when striving to obtain expertise. In this case, before expertise is established in a given area, an individual's performance reaches a stable plateau and no further improvements are observed (a process that is often referred to as premature, or early, automatization). The reasons for this early automatization are poorly understood and may, to some degree, reflect an individual's failure to recognize one or more system factors and/or interactions. This suggests that an important step toward expertise involves metacognition (or thinking about and monitoring one's thinking). Metacognitive processes may help one understand how contextual factors and their interactions can influence the outcome. This effortful step (deliberate "metacognitive" practice), which may mitigate early or premature automatization, may also play a role in expertise whereby individuals employ metacognitive strategies to maintain their expertise. In this phase of expertise development, an individual becomes both the learner and the teacher; that is, he or she becomes highly self-regulated.

Context and Self-regulated (or Self-directed) Learning

Fostering self-directed learning has long been a goal of medical education, particularly as it relates to the use of problem-based learning in medical school.²⁸ In general, definitions of self-directed learning are very similar to what has been termed self-regulated learning in the educational psychology literature.²⁹ However, unlike conceptualizations of self-directed learning, which often consider only cognitive and behavioral factors, many theories of self-regulation have grown from well-established, motivational perspectives and, as such, include consideration for affect (motivation and emotion) and context, as well as cognitive and behavioral factors of the individual. Moreover, theories of self-regulation have a long-standing history of practical relevance to understanding how people function (particularly, how people think and learn). Although several different models of self-regulation exist, social-cognitive theories are unique in their emphasis on contextual features of the setting and how these features interact with personal factors (e.g., beliefs, emotions, and intelligence) and personal behaviors (e.g., help-seeking behaviors, use of deep processing strategies, and use of environmental control strategies) to improve (or hinder) learning and performance. Thus, self-regulated learning can be thought of as an interactive phenomenon, the product of a dynamic interchange between personal and situational influences. As such, self-regulated learning, and self-regulation more generally, is considered by many contemporary psychologists to be a core, defining, and distinctive feature of human activity.³⁰

Research conducted over the last 20 years in the field of education has consistently shown that to achieve personal success (academic and otherwise), individuals must learn how to effectively self-regulate.³¹ In fact, research performed in various educational contexts reveals that successful students possess more than just smarts; they engage in self-generated, goal-directed activities and apply effective strategies to optimally adjust their environment in the service of their goals.³² In doing so, highly self-regulated learners exert appropriate

effort to achieve success, enjoy the challenge of the activity, use appropriate learning strategies, and display high levels of task-specific confidence. In contrast, less-regulated individuals put forth little effort, lose interest in the activity, are unable to apply appropriate strategies, and have low confidence.

Implications for Educational Research

As the preceding discussion indicates, we embrace the notion of context as more than the physical location or additional information surrounding the presenting complaint in the clinical encounter. Instead, we see context in medical education as referring to a complex system that evolves over time. The resulting outcome is driven by interactions and feedback among factors in the environment (patient, physician, setting [referred to as practice or encounter factors], and their interactions); these interactions and feedbacks are not predictable and therefore can be nonlinear. We have addressed a number of theories from fields outside medicine to illustrate how others have conceptualized the notion that individual system components can (and do) interact with other system factors. We have also detailed two theories that we believe incorporate all system components reflected in the clinical encounter.

All the theories that we have discussed in this article view context as more than just the setting or additional content, and the implications of these theories for teaching, learning, and research are numerous. We detail some of these implications in the following paragraphs.

First, we believe researchers should consider studying the factors unique to each participant and those factors' interactions. For example, one can begin with extreme "phenotypes," or prototypical representations of a given condition, in a given system component measure (patient, physician, or encounter factor that constitute the context). Such an approach is consistent with the literature on the development of expertise. It also allows for potential exploration of the magnitude of interactions, consistent with situation awareness.

Sophisticated modeling and analysis techniques should be considered when analyzing component interactions (e.g., nonlinear approaches such as chaos theory as well as hierarchical linear modeling, structural equation modeling, and generalizability). As stated earlier, chaos theory does stipulate that the outcome(s) are bounded (or restricted) and thus are largely predictable or explainable when taking into account the initial conditions (like the pile of sand example). Thus, analyzing complex or chaotic systems involves the use of sophisticated analysis approaches and the recognition that all variance is not likely to be accounted for prospectively.

From a self-regulation perspective, several important questions emerge for research. First, do the self-regulatory competencies of physicians (and physicians in training) change over the course of their medical education, as the context of their education changes? Second, are expert physicians more "self-regulated" than relative novices? And finally, do schools of medicine do anything to help future clinicians develop those self-regulated thoughts, feelings, and actions that would seem to be so critical for effective performance in various contexts? When considered from a self-regulated learning perspective, these become significant practical issues for medical educators.

Self-regulation implications also include consideration of the use of step-wise, “social to self training” in medical education. Such an approach involves several levels of skill acquisition, including an observational level (discriminating the correct form of a skill by watching a capable model), a simulation level (copying the general form of a proficient model), a self-controlled level (practicing a skill in a controlled environment without direct supervision from a model), and a self-regulated level (practicing a skill under dynamic, real-world conditions and modifying that skill based on performance outcomes).^{33,34} A self-regulated learning view also suggests that researchers explore important (and under-researched) physician factors that may influence physician performance. For example, educational psychologists have just recently acknowledged the critically important role that achievement-related emotions play in learning and performance.³⁵ Although limited, recent research findings have indicated that positive emotions (e.g., enjoyment) facilitate the use of flexible, deep processing strategies that result in superior learning and performance. On the other hand, negative emotions (e.g., anxiety and boredom) have been associated with reduced attention and the use of more superficial, shallow processing strategies that result in suboptimal outcomes.³⁶ These findings have important implications for the learning and performance of medical students, residents, and practicing physicians, all of whom experience a wide range of emotions on a daily basis. We believe the field of medical education would benefit from more empirical work that explicitly explores these personal factors and their dynamic relations.

Situated cognition and ecological psychology suggest effective methods that might be used to achieve active, thriving communities of practice in medicine (and medical education). These two theoretical viewpoints would argue for certain, “authentic” methods to promote learning and transfer (e.g., performance in information-rich, authentic settings).¹⁸ In medical education, this could include research that explores new instructional methods (e.g., videotaping participants and using unannounced standardized patients, both of which could shed additional light on individual factors and interactions). Other emerging fields of work include studying metabolic neuroimaging and electrophysiological approaches of physician and/or patient factors³⁷ and how such factors may influence the context of the encounter.

Finally, we believe educators should attempt to unify theoretical constructs to build additional inclusive theory. This would include more uniform language across fields of research and additional venues for the sharing of ideas. As a start, investigators should be encouraged to use grounded practice, or experiments grounded in some theory validated by one or more research traditions. This involves incorporating theories situated in the “environment of study”.¹⁸ Such work could improve the consistency in the use of terms across research fields, build more inter-field theory and practice implications, and encourage new ideas and approaches for important research topics like context in medical education.

References

1. Hodgson CS, Teherani A, Gough HG, Bradley P, Papadakis MA. The relationship between measures of unprofessional behavior during medical school and indices on the California Psychological Inventory. *Acad Med.* 2007;82:S4-7.
2. Papadakis MA, Arnold GK, Blank LL, Holmboe ES, Lipner RS. Performance during internal medicine residency training and subsequent disciplinary action by state licensing boards. *Ann Intern Med.* 2008;148:869-76.
3. Papadakis MA, Hodgson CS, Teherani A, Kohatsu ND. Unprofessional behavior in medical school is associated with subsequent disciplinary action by a state medical board. *Acad Med.* 2004;79:244-9.
4. Greenburg DL, Durning SJ, Cohen DL, Cruess D, Jackson JL. Identifying medical students likely to exhibit poor professionalism and knowledge during internship. *J Gen Intern Med.* 2007;22:1711-7.
5. Fung CC, Relan A, Wilkerson L. Demystifying "learning" in clinical rotations: do immersive patient encounters predict achievement on the clinical performance examination (CPX)? *Acad Med.* 2007;82:S97-100.
6. Hamdy H, Prasad K, Anderson MB, Scherpier A, Williams R, Zwierstra R, et al. BEME systematic review: predictive values of measurements obtained in medical schools and future performance in medical practice. *Med Teach.* 2006;28:103-16.
7. Markert RJ. The relationship of academic measures in medical school to performance after graduation. *Acad Med.* 1993;68: S31-4.
8. Thordarson DB, Ebrahimzadeh E, Sangiorgio SN, Schnall SB, Patzakis MJ. Resident selection and predictors of performance: can we be evidence based? *Clinical Orthop Related Research.* 2007;459:255-9.
9. Durning SJ, Cohen DL, Cruess D, McManigle JM, MacDonald R. Does student promotion committee appearance predict below average performance during internship: a 7 year study. *Teach Learn Med.* 2008;20:267-72.
10. Norman GR, Wenghofer E, Klass D. Predicting doctor performance outcomes of curriculum interventions: problem-based learning and continuing competence. *Med Educ.* 2008;42:794-9.
11. Connell MW, Sheridan K, Gardner H. Developing childhood proclivities into adult competencies: the overlooked multiplier effect. In: Sternberg RJ, Grigorenko EL, eds. *The Psychology of Abilities, Competencies, and Expertise.* New York: Cambridge University Press; 2003:70-93.
12. Dickens WT, Flynn JR. Heritability estimates versus large environmental effects: the IQ paradox resolved. *Psychol Rev.* 2001;108:346-69.
13. Context [definition]. In: Merriam-Webster Online Dictionary [Internet]; c2011. [updated 2009; cited 2010 Jan 21]. Available from: <http://www.merriam-webster.com/dictionary/context>.
14. Gleick J. *Chaos: Making a New Science.* New York: Viking Penguin, Inc; 1987.
15. Thatcher RW. Development as a dynamic system. *Curr Dir Psychol Sci.* 1992;1:189-93.
16. Eve RA, Horsfall S, Lee ME. *Chaos, Complexity, and Sociology: Myths, Models and Theories.* Thousand Oaks (CA): Sage Publications; 1997.
17. Pais, A. The uncertainty relations, with a look back at the correspondence principle. In: *Niels Bohr's Times: In Physics, Philosophy, and Polity.* Oxford (England): Oxford University Press; 1991:304-9.
18. Wilson BG, Myers KM. Situated cognition in theoretical and practical context. In: Jonassen D, Land S, eds. *Theoretical Foundations of Learning Environments.* Mahwah (NJ): Erlbaum; 1999.
19. Young, MF. Instructional design for situated learning. *Educ Tech Research and Development.* 1993;41:43-58.
20. Bredo E. Reconstructing educational psychology: situated cognition and Deweyan pragmatism. *Educ Psychol.* 1994;29:23-35.
21. Gibson JJ. *The Ecological Approach to Visual Perception.* Hillsdale (NJ): Erlbaum; 1986.
22. Endsley MR. Expertise and situation awareness. In: Ericsson KA, Charness N, Feltovich PJ, Hoffman RR, eds. *The Cambridge Handbook of Expertise and Expert Performance.* New York: Cambridge University Press; 2006:633-53.
23. van Merriënboer JJG, Sweller J. Cognitive load theory and complex learning: recent developments and future directions. *Educ Psychol Rev.* 2005;17:147-77.
24. Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ.* 2005;39:98-106.
25. Ericsson KA, Kintsch W. Long term working memory. *Psychol Rev.* 1995;102:211-45.
26. Ericsson KA. Protocol analysis and expert thought: concurrent verbalizations of thinking during experts' performance on representative tasks. In: Ericsson KA, Charness N, Feltovich PJ, Hoffman RR, eds. *The Cambridge Handbook of Expertise and Expert Performance.* New York: Cambridge University Press; 2006:223-41.

27. Ericsson KA. The influence of experience and deliberate practice on the development of superior expert performance. In: Ericsson KA, Charness N, Feltovich PJ, Hoffman RR, eds. *The Cambridge Handbook of Expertise and Expert Performance*. New York: Cambridge University Press; 2006:683-703.
28. Hmelo CE, Evensen DH. Problem-based learning: gaining insights on learning interactions through multiple methods of inquiry. In: Evensen DH, Hmelo CE, eds. *Problem-based Learning: A Research Perspective on Learning Interactions*. Mahwah (NJ): Erlbaum; 2000:1-16.
29. Zimmerman BJ, Lebeau RB. A commentary on self-directed learning. In: Evensen DH, Hmelo CE, eds. *Problem-based learning: A Research Perspective on Learning Interactions*. Mahwah (NJ): Erlbaum; 2000:299-313.
30. Zimmerman BJ. Attaining self-regulation: a social cognitive perspective. In: Boekaerts M, Pintrich PR, Zeidner M, eds. *Handbook of Self-regulation*. San Diego (CA): Academic Press; 2000:13-39.
31. Schunk DH, Zimmerman BJ, eds. *Motivation and Self-regulated Learning: Theory, Research, and Applications*. New York: Erlbaum; 2008.
32. Ertmer PA, Newby TJ. The expert learner: strategic, self-regulated, and reflective. *Instr Sci*. 1996;24:1-24.
33. Zimmerman BJ, Tsikalas KE. Can computer-based learning environments (CBLEs) be used as self-regulatory tools to enhance learning? *Educ Psychol*. 2005;40:267-71.
34. Durning SJ, Artino AR, Holmboe E. On regulation and medical education: sociology, learning, and accountability. *Acad Med*. 2009;84:545-7.
35. Schutz PA, Pekrun R, eds. *Emotion in Education*. San Diego (CA): Academic Press; 2007.
36. Pekrun R. The control-value theory of achievement emotions: assumptions, corollaries, and implications for educational research and practice. *Educ Psychol Rev*. 2006;18:315-41.
37. Krampe RT, Baltes PB. Intelligence as adaptive resource development and resource allocation: a new look through the lenses of SOC and expertise. In: Sternberg RJ, Grigorenko EL, eds. *The Psychology of Abilities, Competencies, and Expertise*. New York: Cambridge University Press; 2003:31-69.

Chapter 3

Aging and Cognitive Performance: Challenges and Implications for Physicians Practicing in the 21st Century

Published in:

The Journal of Continuing Education in the Health Professions 2010;30:153-160

Steven J Durning, Anthony R Artino Jr., Eric Holmboe, Thomas J Beckman,
Cees van der Vleuten, Lambert Schuwirth

Abstract

The demands of physician practice are growing. Some specialties face critical shortages and a significant percentage of physicians are aging. To improve healthcare it is paramount to understand and address challenges, including cognitive issues, facing aging physicians. In this article, we outline several issues related to cognitive performance and potential implications associated with aging. We discuss important findings from other fields and draw parallels to the practice of medicine. In particular, we discuss the possible effects of aging through the lens of situated cognition theory, and we outline the potential impact of aging on expertise, information processing, neurobiology, intelligence, and self-regulated learning. We believe that work done in related fields can provide a better understanding of physician aging and cognition, and thus can inform more effective approaches to continuous professional development and lifelong learning in medicine. We conclude with implications for the healthcare system and areas of future research.

Introduction

The demands of medical practice are growing. Knowledge about diseases and their therapies is rapidly expanding, and the number of specialty journals and articles in the health sciences are emerging at a frantic pace. The quantity of patients seen by providers, acuity of illness for hospital admissions, and complexity of problems in outpatient practice have all increased dramatically over the past decade. Additionally, there has been an explosion of new technologies, such as electronic medical records (EMRs), Internet resources for physicians and patients, and new diagnostic or therapeutic devices. Further compounding the demands of medical practice are the growing number of elderly patients and the critical shortage in primary care clinicians.¹ For example, the United States population is expected to increase to 349 million by 2025, and 1 in 5 adults in the U.S. will be age 65 or older by 2030.² These and other phenomena arguably make the practice of medicine more challenging than in previous decades.

Unless substantial changes to the physician workforce occur, the average age of the remaining physicians in several specialties will also continue to rise. For instance, in the United States, interest in primary care has declined sharply and nearly 1 in 5 physicians who started their careers in primary care are planning to either enter other specialties or go into other careers altogether within 10 years.^{3,4} In Australia, about 25% of the medical workforce is at least 55 years of age, and the percentage of physicians age 65 or above is expected to reach 20% in the next 20 years.⁵

Given these practice realities, older physicians will remain an essential part of the physician workforce. Accordingly, we must understand the implications of aging on cognitive performance, particularly because physicians' medical practices remain largely autonomous. Moreover, consumers who count on medical practitioners to self-regulate their profession, are not medical experts and may be inadequately informed about the quality of their diagnostic and therapeutic outcomes.

It is important to highlight that aging, *per se*, does not necessarily result in cognitive performance impairment. Indeed, a consistent finding from empirical literature outside of medicine is that cognitive performance is more variable as one ages.^{6,7} This variability suggests that important individual differences exist, that aging is just one of several factors that may impact performance,^{6,7} and that performance might be broadly determined by characteristics ranging from intelligence to personality.⁸⁻¹⁰ Additionally, some physician cognitive performance characteristics may actually *improve* with age.¹¹⁻¹² Taken together, the impact of aging on cognitive performance cannot be assumed to affect all physicians equally, and frameworks that can account for complex interactions between factors contributing to cognitive performance are needed.

In this article we will outline some of the issues facing aging physicians' cognitive performance, discuss important findings from fields outside of medicine, and draw parallels to medical practice. To do this, we conducted a thematic review of articles addressing aging and cognitive performance, searching the MEDLINE and ERIC literature to identify articles of relevance and also reviewing the bibliographies of articles identified to address this topic. Furthermore, in this article, we discuss the effects of aging through the lens of situated cognition theory, which places the physician's cognition within the larger physical and social context of interactions, a perspective that can help to explain the potential impact of aging

on expertise, neurobiology, information processing, and self-regulated learning. Importantly, the situated cognition framework does not assume that aging inexorably leads to reduced cognitive performance. Instead it argues that individual factors in the physical and social context of interactions must also be considered.

Situated cognition theory

Situated cognition seeks to place an individual's cognition (thinking) within the larger physical and social context of human interactions. It provides a "person plus" framework, with individual thinking and knowing (or processing) in the setting of social action. In other words, the individual's thoughts and actions are uniquely tied to (and cannot be completely separated from) the specific social situation in which those thoughts and actions occur.^{13,14} Thinking (cognition) occurs in specific social settings, and situated cognition argues that failing to acknowledge the contributions of the setting results in a perspective on thinking that does not fully capture the construct; situated cognition argues that thinking (and acting) are context specific. For example, an individual's thoughts while playing cards or Scrabble, writing a paper, or seeing a patient are fundamentally shaped by their environment. Therefore, thinking is not simply determined by what's stored inside one's brain. This vantage point introduces the importance of the physician's environment to our understanding of cognitive frameworks, such as information processing theory. Information processing theory focuses on what happens inside an individual's head, and as such, it largely downplays the influence of the setting (or environment) and the other participants in the interaction. Situated cognition, on the other hand, puts emphasis on the environment, which, in the case of the clinical encounter, would include patients, practice settings, physicians, and their related interactions (see Appendix 1). In doing so, situated cognition builds upon and expands existing cognitive theories.^{5,15-20}

As shown in Appendix 1, situated cognition places or "situates" the "physician factor" in a social context—underscoring how unique patient, physician, and practice factors can interact in dynamic ways. Situated cognition highlights the notion that each physician, patient, and practice factor has the potential to contribute to the outcome of the clinical encounter. We see advanced physician age as an important "physician factor". Furthermore, by identifying these three factors (physician, patient, and environment/practice), and their potential (and myriad) interactions, situated cognition provides a prospective that may not arise when studying the more classic cognitive theories (outlined below) that address individual patient, physician, or practice/environment factors in isolation (if at all).

In the remainder of this article, we introduce several other theoretical frameworks to address how physician age may impact cognitive performance, as well as how it might interact with patient and practice factors (Appendix 1).

Aging and information processing theory

Information processing theories focus on cognitive structures or mental models inside of one's head (the individual's brain). This class of theories largely emerged in the early days of computer processing, and indeed computers were a driving metaphor.¹⁶ Studies¹⁷⁻²⁰ suggest

that physicians use multiple cognitive models in practice. These cognitive models are often broadly grouped into two types: analytic (actively collecting and analyzing data to make diagnostic decisions, which involves mental effort) and non-analytic (automatic decisions, which are based on prior experiences).

Previous papers^{5,6} that explored this topic from the standpoint of aging have yielded conflicting findings. Some studies have found a strong negative correlation between cognitive performance and age,^{21,22} whereas others²³ have observed a positive correlation between cognitive performance and age. Eva^{6,7} has suggested that analytic processing declines with age whereas non-analytic processing (experience-based and automatic) remains stable. This interpretation by Eva is consistent with studies from multiple fields outside of medicine, which support the idea that as age increases, fluid intelligence or “mental efficiency” decreases, and crystallized intelligence, which is domain-specific and experiential, increases.^{8,24} Eva^{6,7} found that many errors made by older physicians would be expected to correlate with “premature closure”. These errors involve the failure to engage analytic processing and suggest that older physicians are more prone to “go from the gut” (stress prior experience) in the decision-making process. For instance, in a study of ability to assess transfusion decision making, older physicians had the lowest knowledge scores but the greatest confidence in their knowledge.²⁵ This study suggests that older physicians place high value on confidence (due to prior experience), and that such confidence outweighs analytic processing (knowledge scores). If Eva’s hypothesis is correct, then a situated cognition perspective would also suggest that, for example, the additional interaction of practice pressures (for example, shorter appointment length or unfamiliar computer systems) and patient factors (for example, disease acuity, classic versus atypical presentation, and/or suggestion of diagnosis) could further increase the impulse to *not* invoke effortful, time-consuming, analytic processing.

Cognitive load theory,²⁶⁻²⁸ a contemporary information processing theory, deals with the limits of working memory capacity and the benefits of long-term memory and automation; it provides another approach to the question of age and its effects on the physician factor. Cognitive load, or the overall cognitive burden on working memory, is typically divided into intrinsic and extraneous load. Intrinsic cognitive load refers to the number of elements that must be processed simultaneously in working memory and is dependent on both the complexity of the material and the individual’s expertise. Extraneous load is additional cognitive load that is not related to the nature of the task but rather to inefficiencies in how the material is presented. Furthermore, according to cognitive load theory, working memory is very limited in terms of both capacity and duration, whereas the capacity of long-term memory is essentially limitless.

According to cognitive load theory, information held in long-term memory is organized and stored in the form of domain-specific knowledge structures known as schemas (or mental models). Schemas transcend organizing and storing information to free-up working memory capacity in two important ways. First, schemas allow for the storage of large chunks of information (that is, “chunking”), thus limiting the number of possible combinations of information units that would otherwise have to be manipulated in working memory. Therefore, individuals with no prior knowledge or schema(s) for a task (for example, establishing a diagnosis), must consider all possible combinations of information units (for example, considering a broader differential diagnosis). Individuals with appropriate prior

knowledge and schema(s) free up more working memory to accomplish other tasks (for example, the nuances of the proper treatment and its complications for the patient) that would have otherwise been used in creating combinations of information units. Although working memory can hold only a limited number of items at a time, the size and complexity of those items are unlimited.⁸ Therefore, complex schemas consisting of huge arrays of interrelated elements can be held in working memory as a single entity.

The second important way that schemas free up limited working memory capacity is through automation. Automation occurs when, through continued practice, information stored in schemas is processed automatically and without conscious effort. Consequently, a person dealing with previously learned material that has been stored and automated in long-term memory is freed from the processing limitations of working memory. Thus, despite the decrements in working memory capacity that may occur with age, cognitive load theory suggests that schema construction and automation could actually liberate cognitive resources for other purposes.

It follows that despite the decreases in speed and capacity of working memory that appear to occur with aging,^{8,24} older physicians may be able to compensate by using elaborate schemas stored in long-term memory. This compensation would not be expected to apply to new situations, but rather would apply to situations in which older physicians have had prior knowledge and experience. For example, chess studies⁸ suggest the quality of first moves (or first diagnostic hypothesis) is better among older experts, who appear to compensate for declines in search and retrieval speed (processing speed) with more refined knowledge-based processes (automation of schemas stored in long-term memory) to enable move selection.

The situated cognition framework, when combined with cognitive load theory, would suggest that the physician factors of age and memory (both working and long-term) would interact dynamically with patient and encounter factors (Appendix 1). These patient, physician, and encounter factors each uniquely impact cognitive load and, potentially, overall physician performance. For example, even with elaborate schemas, working memory limits may be reached if the extraneous load is too large. Consider the potential impact on cognitive load that might occur when instituting a new electronic medical record system for a physician who has little familiarity with computers. Furthermore, although automation through practice (schema formation) reduces intrinsic load, automation does not mean that the limitations of working memory are never reached. Understanding of the effects of practice and patient factors on cognitive load among aging physicians is an area of inquiry that requires further exploration.

Aging and deliberate practice (expertise)

Developing and maintaining expertise in any field requires extensive, sustained practice of the necessary skills. Multiple fields have shown a clear link between the extended engagement in domain-related activities (or deliberate practice) and the attainment of expert performance.^{9,10} Furthermore, numerous studies have revealed that the most talented individuals in a domain require 10 or more years of experience to become an expert and achieve an international reputation in fields such as chess, sports, music, science, and the arts (the so-

called “10-year rule”¹⁰). Deliberate practice^{9,10} argues that expert performance is acquired gradually, that training tasks must be mastered sequentially, and that training strategies are initially designed by a teacher or coach. Deliberate practice requires effort, immediate and informative feedback about performance, and opportunities for repetition and correction of errors. From the situated cognition framework, deliberate practice involves a physician factor (expertise) with gradual increments of patient and/or practice factors, initially under the supervision of a teacher or coach. By focusing on specific components of an activity for practice, cognitive load is made more manageable.

Challenges to aging physicians in this model include declining peak cognitive performance over time (typically in the fifth decade) regardless of the field studied. Research has shown that more current deliberate practice is required in older adults to *maintain* cognitive performance than in younger adults within the same field.⁸ Moreover, evidence suggests diminishing returns from deliberate practice in older adults and that older adults require more time in between deliberate practice sessions than their younger counterparts.⁸ Importantly, studies of experts in other fields who maintain rigorous deliberate practice activities suggest that expertise can be maintained into the 7th and 8th decades.⁸ This finding has important implications for continuing medical education (CME). For example, one should not assume that a 74-year-old physician is not competent, provided that this physician has demonstrated a complement of deliberate practice activities, satisfactory peer and patient ratings, and acceptable recertification exam scores. An additional challenge in medicine is to determine what *types* of deliberate practice activities should be implemented in a clinical environment that is complex and evolves over time (see Appendix 1). The context of medicine is not so clear-cut as a chess board, where the tasks to be performed and the system itself are more constrained.

Aging and neurobiology

The prevalence of decrements in psychomotor ability with age and reduced speed and accuracy in most cognitive-motor tasks has motivated theories of broad intellectual decline with age, such as *general age-related slowing*.²⁹ For example, visual and hearing acuity decreases throughout adulthood with changes beginning in the second and third decades.³⁰ These decrements have important implications for the aging physician who relies on vision and hearing when conducting histories and physicals. As a general rule, negative age effects are more pronounced if the task requires more complex processing, like recall²⁹, which is an important skill for the practice of medicine. For example, adults in their seventh decade typically need about two times as long to process the same tasks as adults in their 20s. Generally speaking, the evidence from research suggests that “normal” aging reduces the speed and efficiency of cognitive, perceptual, and psychomotor functions.⁸ Proposed mechanisms for this decline include reduced working memory capacity, slowing of retrieval and storage to and from working memory (processing speed), deterioration of neural inter-connectedness³¹, and the inability to ignore irrelevant information.³² A similar concept within the field of medicine is “age associated cognitive decline,” which is memory impairment in the elderly compared with normal young adults.³³

Plasticity refers to the capacity of the brain to change cortical representations (neuro-anatomy) as a function of experience. Plasticity is subject to the impact of age. That is, older adults tend to benefit less from performance-enhancing training programs, need more cognitive support during training, and ultimately attain lower performance after training when compared with younger adults. Further, studies show reductions in cognitive performance with loss of brain tissue from injury such as dementia, stroke, or trauma.³⁴ As these co-morbidities are more prevalent in the aged, and individuals may not be aware of cognitive performance change, screening for these conditions in older physicians may be warranted.

From a situated cognition perspective, reductions in one or more of these physician factors (for example, hearing, vision, and speed of cognition) could be exacerbated (or even mitigated) by relevant patient and/or practice factors in terms of patient care. For example, reduced hearing and vision in an older cardiologist could be compensated for by a wealth of expertise (acquired through deliberate practice) in treating a patient with congestive heart failure (patient factor), provided that the patient presenting with heart failure has symptoms and/or findings that the older cardiologist recognizes with seeing hundreds, if not thousands, of patients presenting to him or her with this disease. Further, impairments in psychomotor abilities with aging could be compensated by less cognitive load through rich schemas. Studying the interactions between these factors is needed to understand the implications of age-related decline within an older physician's practice.

Aging and self-directed and self-regulated learning

Self-directed learning is critical to continuous professional growth and the development of expertise; this seems especially true in a global economy where rapidly changing technology can quickly render knowledge and skills obsolete.³⁵ The healthcare field is no exception. So a vital component of many medical education curricula is helping students develop self-directed, lifelong learning skills.³⁶ Effectively directing one's own learning and cognitive performance requires more than just talent, interest, and good instruction; it also involves planning, motivation, and, especially, self-regulation.³⁷ In other words, effective self-directed learners must implement appropriate self-regulatory strategies when they realize, often based on input from teachers or peers, that they are missing certain knowledge and skills.³⁸ These strategies include an awareness of the knowledge and skills they do and do not possess, an ability to set goals and identify what they need to learn in order to solve the problem at hand, receptiveness to feedback, and a capacity for monitoring and evaluating implemented plans to determine if goals are being attained.³⁶ The complexity of the clinical encounter, as presented in the situated cognition framework (Appendix 1), suggests that self-regulation in medicine may be even more difficult than in other fields, with aging likely exacerbating this intricacy. This notion may partly explain discordant findings from the literature on self-assessment—an essential component of effective self-regulation.³⁹

To date, few researchers have explored the extent to which self-regulation develops, is maintained, and how it may decline over a physician's lifetime. One might expect that cognitive changes in adulthood influence the quality and quantity of self-regulation in older physicians. In fact, several scholars have suggested that there may be important develop-

mental differences in individuals' self-regulation and have encouraged researchers to explore the hypothesis that self-regulated learning might progress developmentally across the lifespan.⁴⁰ In one such study of graduate students' cognitive engagement in engineering and education-based programs, Richardson and Newby⁴¹ found that younger students were more likely to use surface processing strategies (that is, limit cognitive engagement to the bare essentials and use rote memorization to learn) and surface motives (that is, meet requirements minimally), rather than more adaptive, deep processing strategies and motives. In the end, if developmental differences do exist, the question for medical educators becomes: How do we most effectively design CME programs that account for these differences in self-regulation such that CME effectively fosters the continued development of essential self-directed learning skills across the lifespan?

Implications and future research

Current cognitive theories, including situated cognition, are consistent with research that shows physician cognitive performance, on average, declines with age.⁴² Situated cognition argues, however, that reductions in cognitive performance demonstrated in a controlled laboratory setting may not be seen in real-life practice. Instead, to understand the effects of age on cognitive performance truly, we must first explore the unique contributions and interactions of the physician, patient, and practice setting, because performance is tightly bound and fundamentally dependent upon the specific situation. Although cognitive performance, on average, may decline with age, the influence (both valence and magnitude) of patient and practice setting factors on cognitive performance are unknown. We believe this is an essential area for future research. We also know that traditional, passive approaches to CME are ineffective.³⁹ We believe that situated cognition is a useful framework for elucidating the potential impact of patient and practice factors on physician cognition⁴³ and provides a potential framework for future directions in continuous professional development and lifelong learning. The implications of this framework have been used to a limited degree in fields outside of medicine.^{12,43}

Regarding *practice factors* (Appendix 1), rapid changes in healthcare systems and environments require that physicians be highly adaptable. Older physicians may be more prone to errors and inefficiencies in a practice model that emphasizes throughput and 15-20 minute patient encounters. The patient-centered medical home, now being promoted by both physician groups and policy-makers⁴⁴, is a care model that could help ensure physicians have sufficient time to see complex cases and apply evidence-based practice at the point of care. Such models of care include support systems to help older physicians manage cognitive load as well as potentially avoid the overconfidence and premature closure traps that are suggested by findings from information processing theory and research. These models could become a part of more active CME approaches.

Another practice factor that has been neglected is the impact of peer support and mentoring. The majority of care in the United States is delivered in office practices of 5 or fewer physicians, with 20% of physicians working in solo practices.⁴⁴ Because of growing demands for productivity, many of these physicians spend little to no time at their local hospital participating in peer activities or other learning endeavors.⁴⁵ Ericsson and others have high-

lighted the importance of mentors and coaches for acquiring and maintaining expertise through deliberate practice, which is a physician factor. Physicians working alone are therefore at greater risk for declining expertise, particularly as they age when more deliberate practice is likely needed to maintain one's skills.^{42,46} Indeed, some have argued that much of the observed decline in cognitive performance with aging can be attributed to reductions in deliberate practice.⁴⁷ Using the situated cognition approach would highlight the idea that physician factors (that is, aging and less deliberate practice) could combined with one or more practice factors (for example, limited to no learning activities for a practice), resulting in larger-than-expected declines in cognitive performance. Indeed, situated cognition theory would predict that such cognitive performance decrements in two or more factors would be much larger than those that would be expected when considering any one factor alone. Situated cognition theory—by addressing specific patient, physician, and practice factors, as well as their interactions—has important implications for individuals involved in healthcare policy and planning. For instance, aging physicians' cognitive performance could be enhanced by strategies to reduce the burden of these factors, such as practice models that improve physician performance and productivity through technology, physician extenders, and access to point of care learning. Educators should also be mindful that physician learning within practice (practice-based learning and improvement) is making an argument for situated cognition (and learning). Therefore, learning should be tailored to physicians' individual needs.⁴⁸ Situated cognition also supports the stance that continuing education guidelines should reflect the importance of performance, cognition, motivation, and context.⁴⁹ Furthermore, situated cognition (and learning) would argue that planners of continuing professional education must consider the dynamic interplay between personal and situational factors,⁴⁹ incorporate these factors into continuing education curricula, and allow physicians adequate time for reflection within this complex educational model.⁵⁰ Indeed, a key tenet of situated cognition (and learning) is the importance of the concept of practice-based learning and improvement, recently endorsed as an essential component of continuous professional development.⁵¹ Ultimately, educators will need to identify novel teaching strategies and resources for assisting physicians with situated learning.⁴⁸

The evolution of EMRs (Electronic Medical Records; a practice factor) will also require that we account for changes in cognition over time by providing more patient and practice-level data in real time and providing efficient point-of-care learning resources. As older adults are less familiar with new technologies they are, in theory, more prone to performance decrements from the additional cognitive load induced by such technologies. However, if EMRs and other technologies for practice were designed with the aging physician in mind, performance could actually improve. To date, electronic tools to assist clinical reasoning have not been terribly useful.⁵² Additionally, when physicians pursue structured educational activities (practice and physician factors), such as society meetings, it appears that they tend to gravitate toward courses and areas they are already comfortable with, instead of areas where they possess their largest deficits.⁵³ Therefore, more research is needed on how best to deliver knowledge and skills at the point of care and how to help aging physicians recognize and fill their own specific gaps (that is, how to support self-regulation; a physician factor). Although recognizing and addressing personal limitations through self-regulation is critically important, this adaptive "habit" is seemingly lacking in medical practice.

Might the modification of *patient factors* aid the aging physician? Groopman⁵⁴ proposed that patients, as well as physicians, should become more knowledgeable about medical decision making to avoid errors in care. Although this strategy has not yet proven to be effective, enhancing the quality of information for patients on the Internet through peer-reviewed sources, such as WebMD and other reputable educational tools, are steps that could theoretically create more informed consumers. Additionally, educating patients (patient factor) about emerging evidence-based and peer-reviewed treatment guidelines and benchmarks for diseases may improve interactions with their physicians, thus improving the quality of care.

The *physician factor* of intelligence is an important consideration for the aging physician. Evidence supports the notion that different types of intelligence may evolve differently over the years. For instance, although mental efficiency (that is, fluid intelligence) may decline progressively after age 40, an individual's ability to apply contextually appropriate problem-solving methods (that is, crystallized intelligence or domain-specific, experience-based knowledge) can continue to grow throughout the life span. This continued growth is thought to occur because crystallized intelligence includes learned skills and accumulated knowledge.⁵⁵ Indeed, these competing processes—declining fluid intelligence versus increasing crystallized intelligence—and their interactions with emerging and increasingly complex systems of practice may be key issues that determine the successes of aging physicians.

Formal scrutiny of doctors' cognitive abilities may be in store for the future, as experts have called for age-related screening for cognitive impairment and rigorous neuropsychiatric evaluations after lapses in standard of care, regardless of age.⁵⁶ We believe that it would be superior to develop reliable and valid tests, so that capable aging physicians could continue to practice at high levels. Indeed, we found no studies testing objective physician cognitive performance and relating it to physician aging. We believe that aging physicians should be encouraged to attend to their physical health, because studies support that regular exercise and leisure activities decrease risk for dementia,⁵⁷ and improve longevity and independence,^{58,59} with age. Additionally, exercise, diet, sleep, and mental health (physician factors) affect the multi-factor, interacted framework of situated cognition. We also anticipate that attention to physical and emotional health will help mitigate the decrements in sensorimotor and processing-speed changes among aging physicians, thereby optimizing the quality of patient care that is provided throughout a physician's professional life.

Now is the time to promote research regarding how continuous professional development (CPD) should be structured over the course of a physician's career. One program that would lend itself to such research is the maintenance of certification (MOC) program. The MOC is a professional, self-regulatory assessment process with the objectives of assuring the public that physicians are competent and helping physicians improve their cognitive performances.⁶⁰ Indeed, other research has shown that motivation, self-regulation and, ultimately, learning and cognitive performance can be enhanced through well-designed education and training programs that apply the tenets of contemporary learning theory.⁶¹ Using situated cognition as a framework, the MOC program is a logical place to study new approaches to motivation, self-directed assessment, and learning activities built around principles of aging that can inform new directions in CPD. One example is a web-based tool that includes practice-improvement modules that lead a physician through a guided as-

assessment of their practice in specific conditions with explicit attention to the system where the physician provides care. Early evidence suggests this guided process helps physicians to uncover knowledge-performance gaps in care and facilitates changes in practice behavior.^{62,63}

References

1. American College of Physicians. How is a Shortage of Primary Care Physicians Affecting the Quality and Cost of Medical Care? Philadelphia: American College of Physicians; 2008.
2. Wu S, Green A. Projection of Chronic Illness Prevalence and Cost Inflation. Santa Monica (CA): RAND Health; 2000.
3. Lipner RS, Bylsma WH, Arnold GK, Fortna GS, Tooker J, Cassel CK. Who is maintaining certification in internal medicine – and why? a national survey 10 years after initial certification. *Ann Intern Med.* 2006;144:29-36.
4. Hauer K, Durning SJ, Kernan WN, Fagan MJ, O'Sullivan PS, Battistone M, et al. Multicenter study of factors influencing medical students' career choices regarding internal medicine. *JAMA.* 2008;300:1154-64.
5. Collier R. Diagnosing the aging physician. *Can Med Assoc J.* 2008;178:1121-3.
6. Eva K. The aging physician: changes in cognitive processing and their impact on medical practice. *Acad Med.* 2002;77:S1-5.
7. Eva K. Stemming the tide: cognitive aging theories and their implications for continuing education in the health professions. *J Cont Educ Health Prof.* 2003;23:133-40.
8. Krampe RT, Charness N. Aging and expertise. In: Ericsson KA, Charness N, Feltovich PJ, Hoffman RR, eds. *The Cambridge Handbook of Expertise and Expert Performance.* New York: Cambridge University Press; 2006:723-43.
9. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med.* 2004;10 Suppl:S70-81.
10. Ericsson KA. The influence of experience and deliberate practice on the development of superior expert performance. In: Ericsson KA, Charness N, Feltovich PJ, Hoffman RR, eds. *The Cambridge Handbook of Expertise and Expert Performance.* New York: Cambridge University Press; 2006:683-703.
11. Rogers C. Growing old; or older and growing. In: Kirschenbaum H, Henderson VL, eds. *The Carl Rogers Reader.* Boston (MA): Houghton Mifflin; 1989:37-58.
12. Beckman TJ. Lessons learned from a peer review of bedside teaching. *Acad Med.* 2004;79:343-6.
13. Bredo E. Reconstructing educational psychology: situated cognition and Deweyian pragmatism. *Educ Psychol.* 1994;29:23-5.
14. Kirshner J, Whitson JA. *Situated Cognition: Social, Semiotic and Psychological Perspectives.* Mahwah (NJ): Erlbaum; 1997.
15. Eva K. The aging physician: changes in cognitive processing and their impact on medical practice. *Acad Med.* 2002;77:S1-5.
16. Mayer RE. Learners as information processors: legacies and limitations of educational psychology's second metaphor. *Educ Psychol.* 1996;31:151-61.
17. Norman GR. The epistemology of clinical reasoning: perspectives from philosophy, psychology, and neuroscience. *Acad Med.* 2000;75 Suppl:S127-33.
18. Custers EJFM, Regehr G, Norman GR. Mental representations of medical diagnostic knowledge: a review. *Acad Med.* 1996;71:S55-61.
19. Eva K. What every teacher needs to know about clinical reasoning. *Med Educ.* 2005;39:98-106.
20. Moulton CA, Regehr G, Mylopoulos M, MacRae HM. Slowing down when you should: a new model of expert judgment. *Acad Med.* 2007;82:S109-16.
21. Norcini JJ, Lipner RS, Benson JA, Webster GD. An analysis of the knowledge base of practicing internists as measured by the 1980 recertification examination. *Ann Intern Med.* 1985;102:385-9.
22. McAuley RG, Henderson HW. Results of the peer assessment program of the College of Physicians and Surgeons of Ontario. *Can Med Assoc J.* 1984;131:557-61.
23. Hobus PPM, Schmidt HG, Boshuizen HPA, Patel VL. Contextual factors in the activation of first diagnostic hypotheses: expert-novice differences. *Med Educ.* 1987;21:471-6.
24. Salthouse TA. Effects of aging on reasoning. In: In: Ericsson KA, Charness N, Feltovich PJ, Hoffman RR, editors. *The Cambridge Handbook of Expertise and Expert Performance.* New York: Cambridge University Press; 2006:589-605.
25. Salem-Shatz SR, Avorn J, Soumerai SB. Influence of clinical knowledge, organizational context, and practice style on transfusion decision making. *JAMA.* 1990;264:471-5.
26. van Merriënboer JJG, Sweller J. Cognitive load theory and complex learning: recent developments and future directions. *Educ Psychol Rev.* 2005;17:147-77.
27. Sweller J, Chandler P. Why some material is difficult to learn. *Cognition Instruct.* 1994;12:185-233.

28. Kalyuga S, Ayres P, Chandler P, Sweller J. The expertise reversal effect. *Educ Psychol*. 2003;38:23-31.
29. Salthouse TA. *A Theory of Cognitive Aging*. Amsterdam: North Holland; 1985.
30. Baltes PB, Lindenberger U. Emergence of powerful connection between sensory and cognitive functions across the adult life span: a new window to the study of cognitive aging? *Psychol Aging*. 1997;12:12-21.
31. Cerella J. Aging and information processing rates in the elderly. In: Birren JE, Schaie KW. *Handbook of the Psychology of Aging*. 3rd Edition. San Diego (CA): Academic Press; 1990:201-21.
32. Hasher L, Stoltzfus ER, Zacks RT, Rympa B. Age and inhibition. *J Exp Psychol Learn*. 1991;17:163-9.
33. Levy R. Aging-associated cognitive decline. Working party of the International Psychogeriatric Association in collaboration with the World Health Organization. *Int Psychogeriatr*. 1994;6:63-8.
34. Krampe RT, Baltes PB. Intelligence as adaptive resource development and resource allocation: a new look through the lenses of SOC and expertise. In: Sternberg RJ, Grigorenko EL, eds. *The Psychology of Abilities, Competencies, and Expertise*. New York: Cambridge University Press; 2003:31-69.
35. Field J. *Lifelong Learning and the New Educational Order*. 2nd ed. Stoke on Trent (UK): Trentham Books; 2006.
36. Hmelo-Silver CE. Problem-based learning: what and how do students learn? *Educ Psychol Rev*. 2004;16:235-66.
37. Zimmerman BJ. Development and adaptation of expertise: the role of self-regulatory processes and beliefs. In: Ericsson KA, Charness N, Feltovich PJ, Hoffman RR, eds. *The Cambridge Handbook of Expertise and Expert Performance*. New York: Cambridge University Press; 2006:705-22.
38. Ertmer PA, Newby TJ. The expert learner: strategic, self-regulated, and reflective. *Instr Sci*. 1996;24:1-24.
39. Davis DA, Mazmanian PE, Fordis M, Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*. 2006;296:1094-102.
40. Greene JA, Azevedo R. A theoretical review of Winne and Hadwin's model of self-regulated learning: new perspectives and directions. *Rev Educ Res*. 2007;77:334-72.
41. Richardson JC, Newby T. The role of student's cognitive engagement in online learning. *Am J Dist Educ*. 2006;20:23-37.
42. Choudry N, Fletcher R, Soumerai S. Systematic review: the relationship between clinical experience and quality of health care. *Ann Intern Med*. 2005;142:260-73.
43. Durning SJ, Artino AR, Pangaro LN, van der Vleuten CPM, Schuwirth L. Redefining context in the clinical encounter: implications for research and training in medical education. *Acad Med*. 2010;85:894-901.
44. American College of Physicians. *The advanced medical home: a patient-centered, physician guided model of health care*. Philadelphia: American College of Physicians; 2006.
45. Physician hospital relationships: the empty club house [Internet]. Irving (TX): Voluntary Hospitals of America; 2003 Apr [cited 2007 Jan 28]. Available from: www.vha.com/portal/server.pt/gateway/-PTARGS_0_2_18962_0_0_18/phyrelationship_execsum.asp.
46. Holmboe ES, Wang Y, Meehan TP, Tate JP, Ho SY, Starkey KS, et al. Association between maintenance of certification examination scores and quality of care for medicare beneficiaries. *Arch Intern Med*. 2008;168:1396-403.
47. Ericsson KA. How experts attain and maintain superior performance: implications for the enhancement of skilled performance in older individuals. *J of Aging and Physical Activity*. 2000;8:366-72.
48. Lockyer J, Fidler H, de Gara C, Keefe J. Learning to Practice in Canada: The Hidden Curriculum of International Medical Graduates. *J Contin Educ Health Prof*. 2010;30:37-43.
49. Knox AB. Building on abilities. *J Contin Educ Health Prof*. 2003;23:141-5.
50. Cervero RM. Place matters in physician practice and learning. *J Contin Educ Health Prof*. 2003;23:S10-8.
51. Moore DE, Pennington FC. Practice-based learning and improvement. *J Contin Educ Health Prof*. 2003;23:S73-80.
52. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. *N Engl J Med*. 2006;355:2217-25.
53. Eva K, Regehr G. "I'll never play professional football" and other fallacies of self-assessment. *J Contin Educ Health Prof*. 2008;28:14-9.
54. Groopman JE. *How Doctors Think*. New York (NY): Houghton Mifflin Company; 2007.
55. Woolfolk A. *Educational Psychology: Active Learning Edition*. 10th ed. New York: Pearson Education Inc; 2008.
56. LoboPrabhu SM, Molinari VA, Hamilton JD, Lomax JW. The aging physician with cognitive impairment: approaches to oversight, prevention, and remediation. *Am J Geriatric Psychiatry*. 2009;17:445-54.
57. Verghese J, Lipton RB, Katz MJ, Hall CB, Derby CA, Kuslansky G, et al. Leisure activities and the risk of dementia in the elderly. *N Engl J Med*. 2003;348:1208-16.

58. Stessman J, Hammerman-Rozenberg R, Cohen A, Ein-Mor E, Jacobs JM. Physical activity, function, and longevity among the very old. *Arch Int Med*. 2009;169:1476-83.
59. Hakim AA, Petrovitch H, Burchfiel CM, Ross W, Ross GW, Rodriguez BL, et al. Effects of walking on mortality among nonsmoking retired men. *N Engl J Med*. 1998;338:94-9.
60. Holmboe ES, Lynn L, Duffy DF. Improving the quality of care via maintenance of certification and the web: an early status report. *Perspect Biol Med*. 2008;51:71-93.
61. Azevedo R, Cromley JG. Does training on self-regulated learning facilitate students' learning with hypermedia? *J of Educ Psychol*. 2004;96:523-35.
62. Holmboe ES, Meehan TP, Lynn L, Doyle P, Sherwin T, Duffy FD. The ABIM Diabetes Practice Improvement Module: a new method for self assessment. *J Cont Educ Health Prof*. 2006;26:109-19.
63. Duffy FD, Lynn LA, Didura H, Hess B, Caverzagie K, Grosso L, et al. Self assessment of practice performance: development of the ABIM Practice Improvement Module (PIM). *J Contin Educ Health Prof*. 2008;28:39-46.

Chapter 4

Making Use of Contrasting Participants Views of the Same Encounter

Published in:

Medical Education 2010;44:953-961

Steven J Durning, Anthony Artino Jr., John Boulet, CPM Van der Vleuten,
Jeffrey LaRochelle, Bonnie Arze, Lambert Schuwirth

Abstract

Context: The practice of medicine involves many stakeholders (or participant groups such as patients, doctors, and trainees). Based on their respective goals, perceptions and understandings, and what is being measured, these stakeholders may have dramatically different viewpoints of the same event. There are many ways to characterize what occurred in a clinical encounter; these include an oral presentation (faculty perspective), a written note (trainee perspective), and the patient's perspective. In the present study, we employed two established theories as frameworks with the purpose of assessing the extent to which different views of the same clinical encounter (a three-component, Year 2 medical student objective structured clinical examination (OSCE) station) are similar to or different from one another.

Method: We performed univariate comparisons between the individual items on each of the three components of the OSCE: standardized patient (SP) checklist (patient perspective); the post-encounter form (trainee perspective), and the oral presentation rating form (faculty perspective). Confirmatory factor analysis (CFA) of the three-component station was used to assess the fit of the three-factor (three-viewpoint) model. We also compared tercile performance across these three views as a form of extreme groups analysis.

Results: Results from the CFA yielded a measurement model with reasonable fit. Moderate correlations between the three components of the station were observed. Individual trainee performance, as measured by tercile score, varied across components of the station.

Conclusions: Our work builds on research in fields outside of medicine, with results yielding small to moderate correlations between different perspectives (and measurements) of the same event (SP checklist, post-encounter form and oral presentation rating form). We believe obtaining multiple perspectives of the same encounter provides a more valid measure of a student's clinical performance.

Introduction

A clinical encounter involves many stakeholders (participant groups), such as patients, attending physicians, residents, medical students and other healthcare professionals. Based on their respective goals, perceptions and understandings of the encounter, the various stakeholder may have dramatically different viewpoints of the same event. Therefore, including each stakeholder's perspective may add important information about what occurred in a clinical encounter.

Assumptions can be made regarding the 'factual' characterization of a clinical encounter that a written note¹ or oral presentation (OP) represents without direct verification by an independent observer or patient. For example, consultants often review a medical record and then interact with the patient without first discussing the case with the referring doctor. Alternatively, consultants will often evaluate a patient after hearing a brief OP (over the phone or in person) from the referring doctor (or trainee). Finally, clinic and ward attending physicians often listen to a trainee's presentation and critique his or her oral and/or written work without having been present in the room during the trainee's intake of the patient information; however, what is presented orally or in writing by the trainee is thought to represent 'what occurred'. Theoretically, there can be discord between the trainee's presentation and what happened from the patient's perspective. In other words, there may be a disconnect between the patient's perspective of the clinical encounter and what is discussed afterwards in the oral or written presentation (trainee's and attending physician's perspective). This disconnect between the patient's perspective of the clinical encounter and what is discussed afterwards in the oral or written presentation may lead to, for example, delays in the delivery of appropriate treatment.²⁻⁴ Accordingly, knowing which, if any, components of a trainee's OP and written note do not correspond with what actually happened in the encounter (in order to better align these perspectives with the patient's perspective) may assist all stakeholders in the clinical encounter by helping the supervisor to focus on areas that may be discordant. Contexts in which these discordant viewpoints are particularly prevalent represent those in which medical errors are more likely to occur. Medical errors can occur when information collected in the encounter is incorrectly documented or inaccurately passed on verbally to another provider.

Few studies have been conducted that compare different stakeholder views of the same event. Theories of situated cognition and ecological psychology⁵⁻⁷ both suggest that views *emerge* or *evolve* during the encounter. Situated cognition proposes that thinking (cognition) must be viewed as 'situated' within the larger physical and social context of the encounter; it does not occur in a vacuum. This theory also argues for a complex interplay between the participants (social context) and their environment (physical context) because views of participants emerge or evolve as the encounter transpires. Ecological psychology embraces the tenets of situated cognition but approaches the same issues from a slightly different perspective, focusing on the agent (participant) and environment interaction. From an ecological psychology perspective, problem solving is not the product of learners' internal cognitions, but emerges as a result of an *intentionally driven* (goal-driven) participant interacting with a very rich information-containing environment. Both theories would argue that it is not possible to separate the participants' cognitions from the environment in which learning occurs.

Considering these theoretical assumptions, we believe that stakeholders are likely to have quite different viewpoints, based on their respective goals for the encounter, effectivities (abilities) and affordances (opportunities for action), with effectivities and affordances based, in part, on participants knowledge and skill. Moreover, situated cognition theory would argue that by dividing components of the encounter into patient, doctor (physician), and practice factors, the potential complexity of these interactions and how that encounter might evolve can become clearer (Appendix 1). Finally, these theories would suggest that the various stakeholders have similar, but certainly not identical, views of what occurred during an encounter. Part of the reason for these similarities and differences is explained by: (i) the presence of different perspectives and expertise (goals, affordances and effectivities); (ii) the rating of different items (e.g., oral versus written presentation form items), and (iii) the fact that the ‘observer’ of the oral and written note requires an accurate account from the trainee of the event (trainee and patient accounting for another perspective). We were interested in exploring these similarities and differences; we were also interested in differences between viewpoints at the performance extremes. To explore these performance extremes, we looked at concordance between views after dividing performance for each viewpoint into tercile scores.

The overarching purpose of this paper was to apply established theories of situated cognition and ecological psychology as frameworks to assess the extent to which different views of the same encounter are similar to or different from one another. More specifically, we compared patient (using the standardized patient [SP] checklist as a proxy), trainee (written note), and doctor (OP) ratings of the same event. Given the dynamic and complex nature of social interaction, whereby the encounter evolves,⁶⁻⁸ and the notion that the items on the different rating forms are not identical, we expected to find small to moderate correlations between the post-encounter form (PEF) (trainee), the SP checklist (patient), and the OP rating form completed by an attending (doctor). From the same theoretical perspectives, we also predicted that performance across the three-station components for individual trainees would be heterogeneous and that high (or low) performance on one station component would not necessarily be associated with similar performance on the other two components.

Methods

Setting

The F. E. Hebert School of Medicine, within the Uniformed Services University of the Health Sciences (USU), is the United States’ only federal medical school. It matriculates 160-170 medical students annually, and students are exposed to both medicine and the military in their curriculum. It uses a traditional curricular model, whereby students complete two years of classroom work followed by two years of clerkship activities. The former is comprised of one year of basic sciences and one year of pre-clinical (or transition to clerkship) courses; the pre-clinical year is taught largely in a small-group settings. This study involved students at the end of their second or ‘pre-clerkship’ year at this medical school.

Objective structured clinical examinations (OSCEs) provide an ideal investigational milieu in which to study differences in viewpoints on the same event between various participants. Therefore, as part of a six-station OSCE, we developed a novel three-stage station to assess the clinical reasoning skills of Year 2 medical students. As we examine all of our medical students with an OSCE at the end of Year 2 of medical school, we chose this cohort for investigation. In this three-stage station, the trainee first interviews and examines an SP who documents the data gathered using a checklist. Next, the trainee completes a PEF (written note) of the event. Finally, the trainee gives and OP to a faculty member that is rated by that faculty member. Each of the three components of this multi-part station was developed with the goal of evaluating one overarching construct: *clinical reasoning*.

This particular OSCE is a graded event at our institution. The multi-step station used for this study is entitled 'the oral presentation station' and its content of this station (Appendix 2) represented material covered in multiple, Year 2 courses using both lecture and small-group formats.

Encounter (the oral presentation station)

The oral presentation station was composed of three parts. All students proceeded through the station in the same sequence. Each student began this station by interviewing and examining an SP who portrayed a patient with new onset diabetes mellitus (DM; Appendix 2). The SP then completed a checklist concerning data gathering and communication skills (Appendix 3), which required a critique of the student's history taking, physical exam and communication skills using both dichotomous and Likert-scale items. The items on this SP checklist had been vetted by a group of subject matter experts, including inter-departmental leaders with expertise in various clinical fields in medical education.

Next, the student completed a PEF, which required the candidate to answer the following free-text items assessing written communication skill and synthesis of information: summary statement; problem list; differential diagnosis; most likely diagnosis, and supporting data (Appendix 4). This form was graded by an attending physician (SJD) who knew the case parameters but did not listen to the student's OP or view the first step (SP-student interaction) in this study; the possible points per section is shown in Appendix 4. A second attending physician (BA) also reviewed each of the PEF scores to help ensure consistency in scoring. These two attending physicians then met to discuss scoring discrepancies until consensus was reached. The consensus score for the station was used for analysis purposes.

Finally, the student entered a second exam room (third step) where he or she presented the patient to an attending physician. This attending physician rated the student's OP using Likert scale-based items with behavioral anchors on scales of 1 to 5 or 1 to 3, depending on the item. The following components were assessed: pace; history of present illness (HPI); additional history; physical exam; overall organization; summary statement; semantic competence, and overall score (Appendix 5). The attending physician who graded the student's OP did not review the data from the PEF or see the SP encounter. The attending physician was, however, aware of the case parameters (i.e., was given and expected to review the new onset DM case; Appendix 2). Further, students were encouraged to reference the PEF when giving their OP.

Measurements

Each of the evaluation tools used in this study was developed through inter-departmental vetting. This three-component OSCE station was created to evaluate the ability of Year 2 students to accurately arrive at the diagnosis of a patient with new onset DM. All components of this multi-part station were developed with the construct of evaluating clinical reasoning in mind.

The SP checklist was developed to focus on specific behaviors and actions that could be observed in the trainee by a non-doctor and that might be associated with diagnostic ability (e.g. eliciting information on risk factors, associated symptoms and relevant physical findings could help the student make informed decisions regarding the diagnosis). Each of the SPs used in this study went through several hours of training in how to portray the case, as well as in how to score the SP checklist. SPs rated the trainee on the checklist immediately following the encounter with the trainee. Each checklist item had equal weighting (Appendix 3).

The PEF was developed after a review of the clinical reasoning literature by several study authors; it was scored by two of the authors (SJD and BA). The raters met and discussed goals and scoring by section. Follow-up discussions were held after coding approximately a quarter of the PEFs to ensure consistency in the application of the scoring method and to address discrepancies; scores were then revised based on these discussions. A student's score was calculated according to both individual section and overall score. The section scores on the form all referred to free text items and students were asked to construct a summary statement, problem list, differential diagnosis, most likely diagnosis and supporting data for the most likely diagnosis (Appendix 4).

The OP Likert-scale items were created after a review of the literature (conducted by SJD, JLR) and inter-departmental discussion with experts in various specialties in medicine and medical education. Individual OP raters received training on the use of the form and the behavior-related anchors. The OP raters were inter-departmental educators who were responsible for teaching students on the Year 3 clerkships in internal medicine and family medicine. These raters also received and reviewed the script of the case prior to scoring students.

This study was approved by the USU Institutional Review Board.

Analysis

To address the question of different views of the same encounter, we first performed an assessment of the overall measurement model (i.e., the three instruments used to assess the three different views of the encounter) using confirmatory factor analysis (CFA) techniques. Specifically, a CFA was conducted to examine the convergent (measures that should be related are related) and discriminant (measures that should not be related are not related) validity of the multiple measures of the same encounter. In this analysis, maximum likelihood estimation was used to estimate the parameters and a chi-square test was conducted to assess model fit. Generally, a non-significant chi-square result indicates a good model fit.⁹ However, because the chi-square test is influenced by the sample size and the magnitude of the correlations between variables, several additional fit indices were also consid-

ered. These indices included the chi-square: degrees of freedom ratio (also referred to as the normed chi-square), the comparative fit index (CFI), and the root-mean-square error of approximation (RMSEA). According to Hu and Bentler,¹⁰ a well-fitting model should have a normed chi-square less than 3.0, a $CFI \geq 0.90$, and a $RMSEA \leq 0.06$.

The CFA analysis allowed us to model the relationships among the hypothesized constructs using the items in each instrument (i.e., the observed data). Based on theories of situated cognition and ecological psychology, we expected to find evidence of convergent and discriminant validity of the three-factor model, with moderate correlations between the three latent variables that represented the three different views of the encounter. We hypothesized that the latent structure of three instruments would be characterized by three first-order factors representing the three different views of the encounter.

To assess extreme performance trends across this multi-part station, we divided trainee performance into tercile scores to see if performance extremes were consistent across the three performance domains. Although it would seem unlikely that a student could perform well on one component (e.g., SP checklist) and not the others (e.g., PEF), or vice versa, we sought to test this assumption. Descriptive statistical analyses were performed using SPSS Version 12 (SPSS, Inc., Chicago, IL, USA); the CFA was conducted using AMOS 7.0 (SPSS, Inc.).¹¹

Results

Respondents

All students ($n = 170$) completed the oral presentation station. The topic (new-onset DM) in the oral presentation station was explicitly addressed in the Year 2 curriculum as a small-group session in the Introduction to Clinical Reasoning course. The demographics of this class of students are similar to those from other medical schools in the USA. All items on the rating forms were completed with the exception of 21% of trainees left the summary statement on the PEF blank.

Confirmatory Factor Analysis (CFA)

Results from the CFA provided support for the hypothesized, three-factor measurement model (Figure 1 and Table 1). In this analysis, correlations among the 18 observed variables were calculated. The three latent variables were the three factors hypothesized to represent the three different views of the encounter, and the 18 observed variables were the actual items within the three assessment tools. We freely estimated regression weights for 15 of the 18 observed variables (one item per factor served as a marker variable). In addition, we freely estimated the covariance between the three latent constructs.

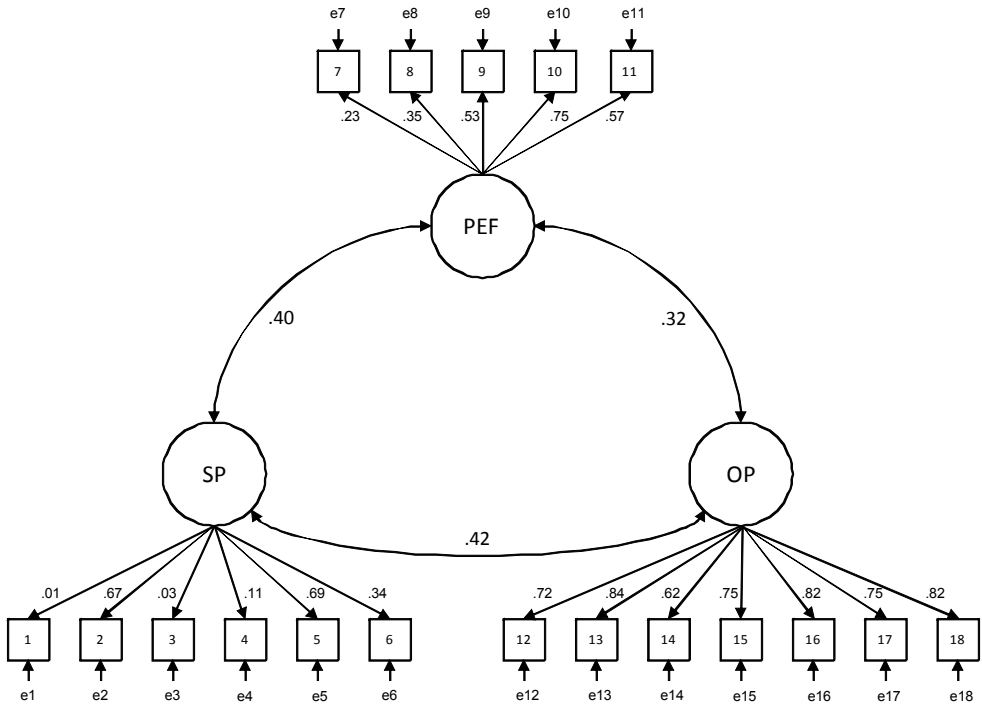


Figure 1: Parameter estimates for the confirmatory factor analysis. Note: item numbers correspond to the item numbers listed in Table 1. The resulting goodness-of-fit indices for the three-factor model suggested a good fit between the model and the data: $\chi^2 (132, n = 170) = 220.50, p < 0.001$, normed chi-square = 1.67, the CFI = 0.90, and the RMSEA = 0.06. Table 1 provides parameter estimates for this three-factor model. Overall, the standardized factor pattern coefficients, which are similar to standardized regression coefficients in multiple regression or factor loadings in exploratory factor analysis, were consistent with the hypothesized model. Specifically, most items had moderate to high standardized pattern coefficients (i.e., greater than or equal to .30) on their corresponding factor. There were, however, four items with relatively low standardized pattern coefficients: three on the SP form (cardinal symptoms, pertinent negatives and risk factors), and one in the PEF form (summary statement). Finally, an internal consistency analysis was conducted on the items retained in each factor and the resulting estimates were deemed reasonable: Cronbach's alpha = 0.64, 0.63, and 0.90 for the SP, PEF and OP scores, respectively.

Table 1: Confirmatory Factor Analysis

	SP	PEF	OP
1. Cardinal symptoms	0.012	–	–
2. Associated symptoms	0.672	–	–
3. Pertinent negatives	0.028	–	–
4. Risk factors	0.111	–	–
5. Physical exam	0.688	–	–
6. Communication	0.336	–	–
7. Summary statement	–	0.228	–
8. Problem list	–	0.348	–
9. Differential diagnosis	–	0.531	–
10. Most likely diagnosis	–	0.747	–
11. Supporting data	–	0.571	–
12. Pace	–	–	0.718
13. History of present illness	–	–	0.835
14. Context	–	–	0.618
15. Physical exam	–	–	0.753
16. Overall organization	–	–	0.823
17. Summary statement	–	–	0.754
18. Semantic competence	–	–	0.822

Standardized Factor Pattern Coefficients (i.e., factor loadings) for the 18-Item, Three-Factor Model and Hierarchical Relationships (SP = standardized patient checklist, PEF = post encounter form, OP = oral presentation rating form)

Relationships between forms:

Post Encounter Form and Standardized Patient checklist = .40

Post Encounter Form and Oral Presentation Rating Form = .32

Oral Presentation Rating Form and Standardized Patient Checklist = .42

Note: Parameter estimates “fixed” to be zero are reported as dashes (–).

Table 2 displays tercile scores for individual trainees with scores at performance extremes (e.g., students who received both a top as well as a bottom tercile performance score on two or more stations). As shown in the table, a considerable percentage of students (31%, $n = 51$), had such performance variation across components of the station. Twenty-six students (15%) scored in the same tercile across stations (8 in the top, 7 in the middle and 11 in bottom tercile). The majority of trainees (54%) obtained scores that crossed only two terciles (e.g. top and middle or middle and bottom).

Table 2: Tercile Performance Across Station Components for Students Receiving Top and Bottom Tercile Scores*

N	SP	PEF	OP
5	3	1	3
4	3	3	1
6	3	1	2
2	3	2	1
5	3	1	1
6	2	1	3
7	2	3	1
4	1	3	2
4	1	2	3
3	1	3	1
4	1	3	3
1	1	1	3

*: N=number of trainees; SP=standard patient checklist, PEF=written note post encounter form, OP=oral presentation checklist score; 3=upper tercile, 1=lower tercile

Discussion

The results from the CFA were consistent with our expectation that, based on theories of situated cognition and ecological psychology, each of the three viewpoints would largely reflect a unidimensional construct and, further, that the three constructs would be significantly correlated but the effects would be moderate. The factor pattern coefficients for the OP items were all statistically significant and the effects were relatively large, suggesting adequate convergent validity of the OP items on the OP form. That is, the OP construct appears to be unidimensional. By contrast, the significant but moderate to low factor pattern coefficients for the PEF summary statement may reflect the fact that about one third of students did not complete this section. Finally, three items from the SP form (cardinal symptoms, risk factors and pertinent negatives) also had low regression weights. We believe this may reflect difficulties inherent in having a non-diagnostician (non-doctor) attempt to rate implied reasoning, which might also enlarge the gap in perspectives given different affordances and effectivities. Too few items or dichotomous scoring by an SP also may have

contributed to this phenomenon. Moreover, this finding may indicate the existence of one or more additional latent constructs, such as thoroughness in data collection without the ability to synthesize presenting information (e.g., dispersed learner).¹² We believe that making the scale more behaviorally anchored, providing further SP frame-of-reference training, and perhaps increasing the number of items in each of these sections would be likely to improve the validity of the SP form.

The results of the CFA also yielded moderate correlations between the three hypothesized latent constructs (PEF, SP and OP; Table 1 and Figure 1). This finding provides support for our theoretic stance—which is consistent with theories of situated cognition and ecological psychology—that different stakeholders have similar but certainly not identical views of what occurs during an encounter. This result probably reflects, in part, the participants' different perspectives and levels of expertise and the fact that items across measurement forms were not identical. Further, some variances in the criteria (and rating) of success for each evaluation instrument may have added to this effect. These shortcomings should be addressed in future research studies.

In short, we believe that academicians should be mindful of the notion that different stakeholders in medicine can (and do) have contrasting views. We were encouraged to find moderate correlations, a result which suggests that it may be beneficial to collect data from more than one viewpoint (for formative and summative evaluation). The recommendation to collect data from multiple viewpoints is consistent with work-based assessment and the program evaluation literature.¹³⁻¹⁷

Tercile scores demonstrated notable scatter in individual trainee performance across our multi-station OSCE. In particular, over 30% of the students scored in the top (upper) and bottom (lower) terciles of the different measures. This finding suggests that each view or station component can add to our overall understanding of trainee performance through triangulation. We are not arguing that any one view represents 'the truth'; instead, we believe each view can inform the others, as all stakeholders have different affordances and effectivities. For example, a trainee could gather the meaningful data by asking numerous questions and performing all the physical examination maneuvers and yet not be able to synthesize the data into a reasonable problem list or differential diagnosis (a dispersed learner). Examining of multiple viewpoints may reveal such discrepancies across stations, which has implications for remediation and research. Likewise, a trainee who asks only a few questions because he or she cuts corners on history and physical examination may actually have superior problem list generation and differential diagnosis skills, which would be reflected in the PEF. Using only the SP checklist would not reveal this difference.

In its entirety, our study suggests that medical educators should not base their assessments of trainee clinical reasoning performance solely on the review of their written or oral presentation or solely on the review of the trainee-patient encounter. Additional useful information for evaluation and feedback can be gleaned by viewing more than one perspective; that is, by triangulating perspectives. This recommendation is consistent with findings from the direct observation literature.^{18,19}

Several limitations should be considered when interpreting our results. Firstly, we only included one year group of trainees, which is a relatively small sample for analysis. Secondly, the validity of the SP checklist and the reliability and validity of the OP checklist scores had not been previously assessed by empirical verification. That said, we believe our

study provides some preliminary reliability (internal consistency) and validity (relationship between measures and model fit) evidence for these scores based on our CFA findings. Thirdly, although SPs and faculty received frame-of-reference training for the exercise, they may have made scoring errors that could have influenced our results. Even with frame-of-reference training and ratings that are behaviorally anchored, one could argue that expert clinicians, as opposed to SPs, represent the optimal raters for evaluating clinical reasoning. Because of limitations in the availability of staff, and our desire to reduce participant anxiety that may have resulted from having a faculty member in the room with the trainee, we used SPs to score one station component using a behaviorally anchored form. We conclude that our approach is feasible. Frame-of-reference training for SPs was provided within 1 hour. In addition, faculty members were able to listen to the OP and score the form in less than 25 minutes per trainee; faculty scoring of the PEF required two faculty members took approximately 10 minutes per trainee. Fourthly, we may have found stronger correlations with a larger sample of trainees or by using more stations to assess similar constructs. However, even if we had more stations on clinical reasoning, context specificity may impact reliability and could possibly attenuate the correlations, thus impacting validity. This limitation should be addressed in future studies. Finally, what happens in a structured setting (i.e., an OSCE) may not reflect what happens in actual practice (i.e., in the clinic or on the wards); performance in a high-stakes examination may be impacted by anxiety or by a Hawthorne effect, whereby participants show good behavior because they are being watched.

In conclusion, our results suggest that different stakeholders have different views of the same clinical event and we believe academicians need to be cognizant of this observable reality. Situated cognition and ecological psychology would argue that ‘true score’ or true variance is a result of each stakeholder’s view and, further, that interactions between views evolve as an encounter evolves. This is a nonlinear approach to measurement, which can exist when multiple parties with different affordances and effectivities are involved in determining performance. We found evidence for this ‘multiple perspectives’ phenomenon in the small to moderate correlations found between individual items and overall rating forms. We advocate faculty development or frame-of-reference training, as well as triangulation of findings between instruments, to address this contrasting viewpoint phenomenon. By training raters on goals and objectives with observation, consistency in ratings can be improved. By triangulating findings, variance caused by the nature of different viewpoints can also be accounted for. In the end, we believe each stakeholder—patient, learner, and teacher—could benefit (and learn) from being explicitly confronted with the contrasting perspectives of the other stakeholders.

References

1. Szeto HC, Coleman RK, Gholami P, Hoffmann BB, Goldstein MK. Accuracy of computerized out-patient diagnoses in a Veterans Affairs general medicine clinic. *Am J Manag Care*. 2002;8:37-43.
2. Griffin SJ, Kinmonth AL, Veltman MW, Gillard S, Grant J, Stewart M. Effect on health-related outcomes of interventions to alter the interaction between patients and practitioners: a systematic review of trials. *Ann Fam Med*. 2004;2:595-608.
3. Ericsson KA. The influence of experience and deliberate practice on the development of superior expert performance. In: Ericsson KA, Charness N, Feltovich PJ, Hoffman RR, eds. *The Cambridge Handbook of Expertise and Expert Performance*. New York (NY): Cambridge University Press; 2006:683-703.
4. Crosskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med*. 2003;78:775-80.
5. Wilson BG, Myers KM. Situated cognition in theoretical and practical context. In: Jonassen D, Land S, eds. *Theoretical Foundations of Learning Environments*. Mahwah (NJ): Erlbaum; 1999:57-88.
6. Young MF. Instructional design for situated learning. *Educ Technol Res Dev*. 1993;41:43-58.
7. Bredo E. Reconstructing educational psychology: situated cognition and Deweyian pragmatism. *Educ Psychol*. 1994;29:23-35.
8. Eve RA, Horsfall S, Lee ME. *Chaos, Complexity and Sociology: Myths, Models and Theories*. Thousand Oaks (CA): Sage Publications; 1997.
9. Kline RB. *Principles and Practice of Structural Equation Modeling*, 2nd ed. New York (NY): Guilford Press; 2005.
10. Hu L, Bentler PM. Cut-off criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model*. 1999;6:1-55.
11. Arbuckle JL. *Amos 7.0 User's Guide*. Chicago (IL): SPSS Inc; 2006.
12. Bordage G. Elaborated knowledge: a key to successful diagnostic thinking. *Acad Med*. 1994;69:883-5.
13. van der Vleuten CPM, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ*. 2005;39:309-17.
14. Durning SJ, Hemmer P, Pangaro LN. The structure of program evaluation: an approach for evaluating a course, clerkship, or components of a residency or fellowship training program. *Teach Learn Med*. 2007;19:308-18.
15. Kassebaum DG. The measurement of outcomes in the assessment of educational program effectiveness. *Acad Med*. 1990;65:293-6.
16. Holzemer WL. A protocol for program evaluation. *J Med Educ*. 1976;51:101-8.
17. Lockyer J. Multisource feedback in the assessment of physician competencies. *J Cont Educ Health Prof*. 2003;23:2-10.
18. Engel GL. The deficiencies of the case presentation as a method of teaching: another approach. *N Engl J Med*. 1971;284:20-4.
19. Holmboe ES. Direct observation. In: Holmboe ES, Hawkins RE, eds. *Practical Guide to the Evaluation of Clinical Competence*. Philadelphia (PA): Elsevier; 2008: 119-29.

Chapter 5

The Feasibility, Reliability, and Validity of a Post-encounter Form for Evaluating Clinical Reasoning

Accepted for publication in:
Medical Teacher, 2011

Steven J Durning, Anthony Artino Jr., John Boulet, Jeffrey LaRochelle,
Cees van der Vleuten, Bonnie Arze, Lambert Schuwirth

Abstract

Background: Developing feasible, reliable and valid methods for the evaluation of clinical reasoning is challenging.

Aim: To explore feasibility, reliability, and validity evidence for a post-encounter form (PEF) assessing clinical reasoning.

Method: A free-text, post-encounter form was used in an Objective Structured Clinical Examination (OSCE) station to assess clinical reasoning for end-of-second-year medical students. Feasibility was assessed by time to complete form. Inter-rater reliability was assessed by kappa. Validity evidence was obtained by comparing scores from individual items on the post-encounter form and other components in this OSCE station (i.e., standardized patient checklist and oral presentation rating form). Additional validity evidence was gathered by comparing scores on this station with other course performance graded events.

Results: Feasibility and estimated reliability were high, and several lines of validity evidence were supported.

Conclusions: The scores from an end-of-second-year, medical school, post-encounter form yielded adequate psychometric properties and can be used for the evaluation of clinical reasoning. Moreover, this form of assessment and its scoring could translate to other venues.

Introduction

Clinical reasoning is the foundation of clinical practice¹—physicians decide on the diagnosis and then act (institute next steps and/or therapy). Educators continue to search for an outcome-based approach for the assessment of clinical reasoning.² Despite the importance of clinical reasoning for optimal patient care, many questions remain about its acquisition and maintenance.^{2,3} For example, the quality and patient safety literature suggest that errors in diagnosis and management are the most common and pervasive errors in practice, with great cost to society each year.⁴ Additionally, clinical reasoning appears to be dependent upon the clinical presentation of the condition, and expertise in evaluating and treating one condition does not necessarily translate into the proper evaluation and treatment of other conditions (context specificity).⁵

Several studies suggest that helping trainees establish the intermediate steps to the diagnosis may improve diagnostic accuracy.^{3,6-9} Further, the literature from other fields has revealed that improved problem representation (an intermediate step to the solution) is associated with expertise.^{3,6-9} An example of an intermediate step is an encapsulation.^{10,11} Encapsulations are the synthesis of presenting history, physical, laboratory and/or other data into more meaningful abstract terms proximal to the diagnosis, such as recognizing hypotension and tachycardia as “shock”. Other examples of intermediate steps are semantic qualifiers or competence,¹² which refer to abstract, often binary, terms like “acute monoarticular arthritis” or “chronic unilateral edema”.¹²

Post-encounter, written exercises (post-encounter forms or PEFs) are often used as part of Objective Structured Clinical Exams (OSCEs). These written exercises often take the form of multiple-choice questions and, at times, a request for next steps in diagnosis or management. As such, PEFs described in the literature implicitly measure clinical reasoning. These forms, however, have not been used to directly measure clinical reasoning by asking trainees to list the intermediate steps to the diagnosis (summary statement, problem list, differential diagnosis) or by explicitly scoring trainees for proper and improper use of semantic competence and encapsulations. Further, currently described PEFs have not asked trainees to provide additional supporting data for their leading diagnosis, all in free-text format.

The purpose of this paper was to provide evidence to support the feasibility, reliability and validity of a relatively novel free-text, post-encounter form for assessing clinical reasoning in end-of-second-year medical students. Such a free-text, PEF, which includes scoring of intermediate steps to the diagnosis, has not been studied previously. Establishing the psychometric properties of such an instrument could assist medical educators responsible for evaluating clinical reasoning on written evaluation forms in clinical practice.

We hypothesized that with sufficient training, raters should be able to score the PEF consistently. We also hypothesized that trainees with higher semantic competence and encapsulation scores would perform better on a checklist of the encounter completed by a standardized patient (SP), as well as the subsequent oral presentation (OP) of the encounter to a faculty member. We also expected that students with higher PEF scores would perform better on an end-of-course essay exam and an end-of-course shelf exam assessing clinical diagnosis skills.

Method

Setting

The F. E. Hebert School of Medicine, within the Uniformed Services University of the Health Sciences (USU), is the nation's only federal medical school. We matriculate 160-170 medical students each year, and students are exposed to both medicine and military in their curriculum. Students complete two years of classroom work followed by two years of clerkship activities. The former is comprised of 1 year of basic sciences and 1 year of pre-clinical or transition to clerkship courses, with the pre-clinical year being taught largely in a small-group setting. This study involved students in the second or "pre-clinical year" at our medical school. This participant group was chosen as they complete this three component OSCE station as a compulsory component of the second year medical student curriculum.

The encounter used in this study was a multi-step OSCE station for end-of-second-year medical students. This OSCE is a graded event at our institution. The multi-step station used for this study is entitled "the oral presentation station," and the content of the station represented material covered in multiple, second-year courses in both lecture and small-group format.

Measurements

The oral presentation station

Each component of the 3-part oral presentation station was developed through inter-departmental vetting. This three-component OSCE station was created to evaluate the ability of second-year students to accurately arrive at the diagnosis of a patient with new onset diabetes mellitus. Each component of this multi-part station was developed with the intent of evaluating the construct of *clinical reasoning*. The SP checklist was developed to focus on behaviors that could be detected in the trainee by a non-physician, which may be associated with diagnostic ability. For example, eliciting risk factors, associated symptoms, and/or pertinent physical examination findings could allow the student to make more informed decisions concerning the diagnosis. Each of the SPs used in this study went through several hours of training on how to portray the case as well as how to score the SP checklist.

The oral presentation station was composed of three parts, and all students proceeded through the station in the same sequence. Students began this station by interviewing and examining a SP who portrayed new onset diabetes mellitus (Appendix 2). The SP then completed a checklist of data-gathering and communication skills, which required a critique of the student's history, physical exam and communication skills using both dichotomous and Likert-scale items. The items on this SP checklist were selected after vetting by a group of eight subject matter experts, including inter-departmental leaders with expertise in various clinical fields and medical education (Appendix 3).

Next, students completed a PEF, which asked them to answer the following free-text items assessing written communication skill and synthesis of information: summary statement (1 point), problem list (3 points), differential diagnosis (3 points), most likely diagnosis (1 point), and supporting data (2 points; see Appendix 4). This form was graded by an

attending (SJD) who knew the case parameters but did not listen to the student's oral presentation or view the first step (SP and student interaction) in the study. This attending also noted proper and improper use of semantic qualifiers and encapsulations, which we combined and will call the total semantic competence (SC) score for the remainder of this article (1 point for each properly used term and -1 point for each improperly used term). Our classification of semantic competence and encapsulations for PEF utterances is consistent with the literature.¹⁰⁻¹² A total semantic competence score, as well as number of correct (positive) and incorrect (negative) semantic competence, were calculated for each PEF.

After identifying each SC, which the attending underlined on the form, a score was generated as being correct (SC positive) or incorrect (SC negative). The SC scores (positive, negative, and total [i.e., positive minus negative]) were not part of the score that the trainee received for the post-encounter. A second attending (BA) also reviewed each PEF and looked for SCs in the fashion described above. In addition, the second attending (BA) reviewed the PEF section and total scores to ensure consistency in scoring. The two attendings then met to discuss discrepancies in scoring (both scores for grading as well as ungraded semantic competence/encapsulation totals) until consensus was reached; student scores were based on consensus.

Finally, students entered a second exam room (third step) where they presented the patient to an attending. This attending rated the student's oral presentation using Likert-scale items with behavioral anchors on a 1 to 5 scale, with the exception of pace, which was assessed on a 1 to 3 scale. The following components were assessed: pace, history of present illness (HPI), additional history, physical exam, overall organization, summary statement, semantic competence and overall score (see Appendix 5). The overall score was calculated by summing scores on the other items; the maximum possible score was 33. Despite their importance in clinical practice, oral presentations are currently not a part of national examinations on the medical school level and few residency-training programs use such examinations for board certification. The attending who graded the student's oral presentation did not review the data from the PEF or see the SP encounter. This attending was, however, aware of the case parameters (i.e., the attending was given and expected to review the new onset diabetes mellitus case). Further, students were encouraged to use the PEF when giving their oral presentation.

Oral Presentation (OP) Station Instrument Development

The oral presentation PEF was developed after a review of the clinical reasoning literature by several study authors; it was scored by two of the study authors (SJD and BA). The raters met and discussed goals, scoring by section, and semantic competence and encapsulation scoring. Follow-up discussions were held after coding approximately one quarter of the PEFs to ensure consistency in application of scoring method and to address discrepancies; scores were then revised based on these discussions.

The OP Likert-scale items were created after a review of the literature (SJD, JL) and inter-departmental discussion with experts in various specialties in medicine and medical education. Individual oral presentation raters received training on the use of the form and the behavior-related anchors. The oral presentation raters were interdepartmental educators who are responsible for teaching students on the third year clerkship in internal medi-

cine and family medicine. These raters also received and reviewed the script of the case prior to scoring students.

NBME Shelf Examination Score

At the end of the second year of medical school at USU, students take a shelf examination entitled “Introduction to Clinical Diagnosis”. This is a multiple-choice examination that emphasizes the establishment of a diagnosis from a clinical vignette. The examination covers a variety of diagnoses that are often seen by students beginning in the clerkships. The examination is reported as a national mean of 500 with a standard deviation of 100.

ICR Examination Score

In the Introduction to Clinical Reasoning (ICR) Course, students take two multiple-choice examinations and complete one essay examination. The multiple-choice examinations are vignettes followed by most likely, least likely, or next step lead-ins. At the end of the second year of medical school, USU students complete an essay examination pertaining to topics covered in the ICR course. This essay examination was scored by two of the study authors (SJD and BA) and one of the three essay questions pertained to diabetes mellitus.

ICR Small Group Grades

Each student receives a grade for the small group component in the course based on their preparation and participation in each of the 30 small groups offered in the course. This study was approved by the USU Institutional Review Board.

Analysis

Feasibility of the PEF was estimated by student completion rate, faculty time required for scoring the form, and faculty satisfaction with the rating instrument. Inter-rater reliability was calculated as agreement between two separate raters (SJD and BA) on both section scores and semantic competence and encapsulation score. Students’ notes were double scored on all measures.

Validity was assessed in several different ways. Content validity was assessed by discussions with post-encounter form raters and meetings with eight inter-departmental experts to discuss and revise the form. Concurrent validity was assessed by comparing individual scores on the PEF with individual SP checklist ratings (proxy for patient) and OP checklist ratings. We also estimated concurrent validity by comparing PEF scores with an end-of-course shelf examination (Introduction to Clinical Diagnosis) and end-of-course essay exam score in a different case of diabetes. Pearson was used for correlations between forms; that is, to see if the items on each form that rated like items were correlated. We used a p value of <0.01 for statistical significance given that we made multiple comparisons. We converted all scores to percentage points available, given that some of the items had different numeric response scales (i.e., 1-3 versus 1-5). Predictive validity was estimated by assessing the relations between PEF scores and internal medicine clinical performance ratings

and internal medicine NBME shelf examination score. Pearson was used for assessing associations. All analyses were conducted on SPSS version 12 (Chicago, IL, USA).

Results

Respondents

All students ($n = 170$) completed the oral presentation station. The topic (new onset diabetes mellitus) used in the oral presentation station was explicitly addressed in the second-year curriculum as a small-group session in the ICR course. The demographics of this class of students were similar to those from other medical schools in the US. All respondents completed all measurements. One item on the PEF (the summary statement) was not answered by 21% of study subjects.

Feasibility, Reliability and Validity of the PEF

Feasibility for the PEF was estimated to be high. With the exception of the first question on the form (problem list), all students completed the form within the time allotted for the station (12 minutes). Further, faculty raters (SJD and BA) estimated that scoring each PEF took 1-3 minutes, once discussion and consensus on ratings were reached. Faculty from several departments expressed high satisfaction with this rating instrument, stating that it appeared to encompass what is expected in the clinical clerkships and acknowledging that it reflected essential components of the clinical reasoning process. For example, some of these faculty voiced the recommendation that the weight of this PEF should be doubled for calculation of final grade.

Inter-rater reliability was high for both total scores ($\kappa=0.82$) and semantic competence and encapsulation scores ($\kappa=0.92$). The two rater scores were independent; all discrepancies between the two PEF raters were rectified with discussion until consensus was reached. The consensus score was used in all of our analyses.

Content validity for this instrument was supported since the components on the form are consistent with the literature on clinical reasoning skill^{3,6-9} and underscore what is expected of physicians writing notes in clinical practice. Pearson correlations between all measures gathered in the diabetes station are shown in Table 1. Generally speaking, small to moderate correlations (between like categories) were found between scores from the rating instruments. For example, overall performance on the three instruments—SP checklist, PEF, and OP—yielded small correlations (range 0.2-0.3).

Table 1: PEF Rating items and Pearson Correlations (statistically significant ($p<.01$) R values are shown between instruments)

PEF items [^]	OP form items ^{^^}	SP checklist items ⁺
Summary Statement		Associated Symptoms (0.16) Overall Score (0.18)
Problem List		
Differential Diagnosis		
Most Likely Diagnosis	Pace (0.21) HPI (0.28) Overall Organization (0.26) Context (0.19) Physical Exam (0.17) Summary Statement (0.27) Semantic Competence (0.27) Overall Score (0.23)	Associated Symptoms (0.26) Physical Exam (0.25) Overall Score (0.28)
Supporting Data	HPI (0.21) Summary Statement (0.19) Semantic Competence (0.23) Overall Score (0.22)	Associated Symptoms (0.17)
Total Score	HPI (0.23) Overall Score (0.17) Semantic Competence (0.17)	Associated Symptoms (0.22) Overall Score (0.20)
SC Positive		
SC Negative	Overall Organization (-0.30) Semantic Competence (-0.26) Overall Score (-0.20)	Risk Factors (-0.34) Communication (-0.22)
SC Total	Semantic Competence (0.18)	

[^]=PEF or post-encounter form (Appendix 4); SC=semantic competence
^{^^}=OP form or oral presentation form (Appendix 5); HPI=History of Present Illness
⁺=SP checklist or standardized patient checklist (Appendix 3)

Validity evidence for the examination was supported by correlations with concurrent performance measures (Table 2; small group grades for the course, other course examinations, and the NBME Introduction to Clinical Diagnosis shelf examination).

Table 2: Pearson Correlations*

	Small Group Grades	Course Exams	NBME	PEF
PEF	0.30	0.38	0.23	
Small Group Grades		0.77	0.46	0.30
Course Exams	0.77		0.80	0.23

*=All correlations are statistically significant at the .01 level

NBME= National Board of Medical Examiners subject exam for Introduction to Clinical Diagnosis.

Small group grades=overall average score received for teacher rated small group performance using a case based teaching approach

Course exams=two multiple choice exams and one short answer essay exam

PEF=post-encounter form

Discussion

In the present study, employing the oral presentation PEF was feasible. As such, this form may be useful for other institutions seeking to evaluate clinical reasoning on an OSCE station in the preclinical setting. Moreover, students and faculty alike believed that the form reflects essential components of clinical reasoning and students were able to complete the station within the allotted time. The reason for the lower percentage of individuals completing the summary statement question is not known but may reflect the structure of the form, unfamiliarity with problem list construction, or confusion over directions.

The PEF scores showed evidence of good inter-rater reliability as assessed by inter-rater measures for each section, as well as the semantic competence/encapsulation ratings. The high inter-rater reliability is noteworthy considering that relatively little training and discussion between investigators was required to reach consensus. A high internal consistency reliability was also observed on this free-text instrument.

In terms of validity evidence, there were several notable correlations on univariate analysis that are worthy of additional comment (Table 1). Overall, the correlation between PEF summary statement and SP checklist supports the literature on problem representation^{3,6-9}—students who were able to successfully extract key features of the encounter performed better on the diagnostic task of the station as rated by the patient. This also suggests that patients may be able to rate trainee diagnostic abilities if given training as well as specific behaviors and/or actions to rate, even without in depth knowledge of diagnosis. This may be particularly important for institutions that are considering the use of other educators to rate medical student and/or physician performance (e.g., physician assistants, SPs, nurses).

The PEF item, most likely diagnosis, had small to moderate correlations with all OP rating scale items. This finding was not surprising, as the exercise was one of establishing the diagnosis of diabetes mellitus; this finding supports the validity of the PEF. This result also suggests that faculty place heavy weight on the trainee's ability to establish the diagnosis with listening to an oral presentation. Indeed, the ability to collect and convey important history, physical examination, laboratory and other findings in a highly competent fashion (e.g., to be a Reporter) does involve interpretation or ability to formulate the diagnosis.¹³ The most likely diagnosis item on the PEF also correlated with SP associated symptoms,

which is consistent with the literature on problem representation—eliciting associated symptoms requires thinking diagnostically while taking the history. Likewise, PEF item, most likely diagnosis, correlated with the SP physical examination, which suggests that performing selected/pertinent physical exam items requires some understanding of the likely diagnosis. Our findings support the idea that knowledge of the most likely diagnosis is associated with eliciting these key features at the bedside by an independent rater. The PEF item, supporting data, correlated with several items on the SP checklist and oral presentation form (Table 1). Again, this would be expected on this station that emphasized establishing a diagnosis, and this finding provides additional validity evidence for the PEF ratings. Likewise, as expected, overall score on the PEF correlated with overall score on the OP rating form, again adding validity evidence to our PEF.

The lack of association between two PEF items, problem list and differential diagnosis, and the OP rating form and SP checklist were noteworthy. It appears that problem list and differential diagnosis dimensions on the PEF may be assessing different constructs than those reflected on the SP and OP checklists. The construct measured by the PEF may represent early problem representation or knowledge organization,¹² which is consistent with correlations of these items with semantic competence/encapsulation scores.

The results from the semantic competence and encapsulation ratings on the PEF also revealed several notable findings. Negative score on semantic competence/encapsulations (one or more improperly used terms) negatively correlated with overall organization, semantic competence and overall competence on the OP checklist. This suggests that proper use of terminology is important to OP raters who were different than the PEF rater. The correlations between PEF semantic competence/encapsulation scores and OP semantic competence is also notable, suggesting that there may be overlap with use of oral and written medical terms and also adds validity evidence for our PEF. Further, negative scores on semantic competence/encapsulations were also correlated with SP communication scores. This suggests that trainees who do not use medical terminology properly may also have more difficulty than others in communicating questions and replies to patient's queries in lay terms. Adding to this notion, negative SC/encapsulations negatively correlated with risk factors, which is diagnostically driven knowledge/understanding questions on diabetes.

Our results suggest the PEF has adequate content and concurrent validity. The majority of the univariate associations in Table 1 were small to moderate. This finding is consistent with the idea that the rating items were similar but not identical and also reflects the notion that different stakeholders may have different views of the same event. This latter concept is consistent with established educational literature regarding situated cognition and ecological psychology theories.¹⁴⁻¹⁶

It was noteworthy that a form completed in less than 15 minutes demonstrated moderate correlations with in-house examinations in the course, which included an essay examination as well as a national board examination on a variety of clinical reasoning topics. Likewise, small-group grades, reflecting multiple preceptors' evaluations of a student's clinical reasoning, showed moderate correlation with the PEF. We believe that our findings support the use of this form for OSCE stations assessing clinical reasoning.

Our study has several important limitations that should be considered when interpreting our results. First, we included a relatively small sample of trainees from one year. Second, the PEF represents only one station in this multi-station OSCE. Third, although SPs

and faculty received frame-of-reference training for the exercise, they may have made scoring errors. Fourth, the association with shelf examination scores, which are multiple-choice in format, is of limited value. Fifth, we did not gather psychometric evidence to support the use of this form for constructs other than clinical reasoning. We found evidence to support the use of this form for second year medical students. Future investigations could explore the use of this form to assess clinical reasoning for trainees at other educational levels.

References

1. Higgs J, Jones MA, Loftus S, Christensen N, eds. *Clinical Reasoning in the Health Professions*. 3rd ed. Oxford: Butterworth Heinemann; 2008.
2. Schuwirth L. Is assessment of clinical reasoning still the Holy Grail? *Med Educ*. 2009;43:298-9.
3. Ericsson KA. The influence of experience and deliberate practice on the development of superior expert performance. In: Ericsson KA, Charness N, Feltovich PJ, Hoffman RR, eds. *The Cambridge Handbook of Expertise and Expert Performance*. New York: Cambridge University Press; 2006:683-703.
4. Crosskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Acad Med*. 2003;78:775-80.
5. Eva K. What every teacher needs to know about clinical reasoning. *Med Educ*. 2005;39:98-106.
6. Charness N, Tuffiash MI, Krampe R, et al. The role of deliberate practice in chess expertise. *Appl Cognitive Psych*. 2005;19:151-165.
7. Chi MTH, Glaser R, Farr MJ, eds. *The Nature of Expertise*. Hillsdale (NJ): Erlbaum; 1988.
8. Hoffmann RR, Lintern G. Eliciting and representing the knowledge of experts. In: Ericsson KA, Charness C, Feltovich PJ, Hoffman RR, eds. *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge (NY): Cambridge University Press; 2007:203-22.
9. Chi MTH, Glaser R, Rees E. Expertise in problem solving. In: Sternberg RJ, ed. *Advances in the Psychology of human intelligence*. Hillsdale (NJ): Erlbaum; 1982:7-76.
10. Schmidt HG, Boshuizen HPA, Hobus PPM. Transitory stages in the development of medical expertise: the "intermediate effect" in clinical case representation studies. In: *Proceedings of the 10th Annual Conference of the Cognitive Science Society*. Montreal (Can): Erlbaum; 1988:138-45.
11. Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise: theory and implications. *Acad Med*. 1990;65:611-22.
12. Bordage G. Elaborated knowledge: a key to successful diagnostic thinking. *Acad Med*. 1994;69:883-5.
13. Pangaro LN. A new vocabulary and other innovations for improving descriptive in-training evaluations. *Acad Med*. 1999;74:1203-7.
14. Wilson BG, Myers KM. Situated cognition in theoretical and practical context. In: Jonassen D, Land S, eds. *Theoretical Foundations of Learning Environments*. Mahwah (NJ): Erlbaum; 1999:57-88.
15. Young MF. Instructional design for situated learning. *Educ Tech Research and Development*. 1993;41:43-58.
16. Bredo E. Reconstructing educational psychology: situated cognition and Deweyian pragmatism. *Educ Psychol*. 1994;29:23-35.

Chapter 6

The Impact of Selected Contextual Factors on Experts' Clinical Reasoning Performance

Accepted for publication in:

Advances in Health Science Education, 2011

Steven J Durning, Anthony R Artino Jr., John R Boulet, Kevin Dorrance,
Cees van der Vleuten, Lambert Schuwirth

Abstract

Background: Context specificity, or the variation in a participant's performance from one case, or situation, to the next, is a recognized problem in medical education. However, studies have not explored the potential reasons for context specificity in experts using the lens of situated cognition and cognitive load theory (CLT). Using these theories, we explored the influence of selected contextual factors on clinical reasoning performance in internal medicine experts.

Method: We constructed and validated a series of videotapes portraying different chief complaints for three common diagnoses seen in internal medicine. Using the situated cognition framework, we modified selected contextual factors—patient, encounter, and/or physician—in each videotape. Following each videotape, participants completed a post-encounter form (PEF) and a think-aloud protocol. A survey estimating recent exposure from their practice to the correct videotape diagnoses was also completed. The time given to complete the PEF was randomly varied with each videotape. Qualitative utterances from the think-aloud procedure were converted to numeric measures of cognitive load. Survey and cognitive load measures were correlated with PEF performance. Pearson correlations were used to assess relations between the independent variables (cognitive load, survey of experience, contextual factors modified) and PEF performance. To further explore context specificity, analysis of covariance (ANCOVA) was used to assess differences in PEF scores, by diagnosis, after controlling for time.

Results: Low correlations between PEF sections, both across diagnoses and within each diagnosis, were observed (r values ranged from -0.63 to 0.60). Limiting the time to complete the PEF impacted PEF performance ($r=0.2$ to 0.4). Context specificity was further substantiated by demonstrating significant differences on most PEF section scores with a diagnosis (ANCOVA). Cognitive load measures were negatively correlated with PEF scores. The presence of selected contextual factors appeared to influence diagnostic more than therapeutic reasoning ($r= -0.2$ to -0.38).

Conclusions: Contextual factors appear to impact expert physician performance. The impact observed is consistent with situated cognition and CLT's predictions. These findings have potential implications for educational theory and clinical practice.

Introduction

Clinical reasoning involves both establishing the diagnosis (diagnostic reasoning) and deciding on a plan of action that is specific to the patient's circumstances and preferences (therapeutic reasoning). There are many challenges to evaluating reasoning in the clinical setting. For example, multiple studies have found that performance across situations (cases) can vary greatly—this is often referred to as the problem of context specificity.^{1,2} In other words, context specificity is the phenomenon invoked to explain the low correlations in performance across different clinical problems for a given physician or medical student. For example, Norman et al.¹ previously demonstrated that for residents and medical students presented with a series of simulated patient problems, low correlations were found across problems with the exact same diagnosis and a different chief complaint. Indeed, the name “context specificity” was given to this phenomenon to serve as a reminder that content is an insufficient explanation for this problem.^{2,3} Further, some researchers have remarked that context specificity is the “one truth in medical education research”.⁴ We sought to explore the components of context that might lead to this phenomenon.

Context defined and theoretical frameworks

Context can be defined as “the weaving together” of the participants and the setting in the clinical encounter.^{5,6} This weaving together, or interaction, creates meaning that evolves during the encounter. Thus, context is more than the individual participants in the clinical encounter or the setting where the encounter occurs—it entails the physician, the patient, the setting, and the interactions of all three.⁶

Situated cognition is a contemporary educational theory; it argues that context emerges from participant and setting interactions. In this theoretical framework (Appendix 1), context can be defined as the weaving together of participants (patient, physician and trainees, and/or family, as applicable), within a setting, to create meaning which evolves through interactions. Furthermore, in this framework, the outcome of the encounter is based on multiple interacting inputs (participants and setting). Thus, situated cognition provides a unique lens through which to view the problem of context specificity.

Cognitive load theory (CLT) is a contemporary information processing theory that addresses the limits of working memory capacity and its relationship to instructional interventions. CLT is one of the leading theories for investigations into cognitive processes.⁷⁻⁹ A tenet of CLT is element interactivity. Although the information that physicians must process varies on many dimensions, the extent to which relevant informational elements interact has been found to be a critical feature. Information varies on a continuum from low to high in element interactivity. Each piece of low-element interactivity information can be processed and understood individually without consideration of any other information elements. Each piece of high-element interactivity information requires consideration of other elements to be processed and understood.⁷⁻⁹ As such, high-element interactivity information results in greater cognitive load; that is, it places greater demands on cognitive processing.⁸ From the situated cognition framework's perspective (Appendix 1), the multiple contextual factors inherent to a clinical encounter, by virtue of their interactions and the emergence of meaning, would be predicted to have high element interactivity and thus

would be expected to lead to high cognitive load. Further, these interactions between participants lead to emerging goals, and CLT would predict that emerging goals (goal-free to goal-specific principle) would increase cognitive load for the participants in the encounter.⁷⁻⁹

Semantic qualifiers are abstract, often binary, terms used to characterize or classify a patient's presentation.¹⁰ Examples of semantic qualifiers include acute, bilateral, and chronic. The presence of semantic qualifiers is readily measured on think-aloud analysis and thus was included in our study. It is believed that such qualifiers may help physicians associate or organize knowledge in memory, and one expected manifestation of this, according to CLT, would be decreased information element interactivity. Some studies have suggested that the use of these terms may be associated with improved diagnostic performance.¹¹ The improper (or incorrect) use of semantic qualifiers in experts who are given the task to diagnose and treat problems familiar to their practice therefore could be considered as a manifestation of high cognitive load. Disfluencies, such as articulation errors, deletions, and substitutions that are akin to incorrect use of semantic qualifiers, as well as increased pauses, have been associated with high cognitive load in other fields.^{12,13}

Studies have not explicitly explored the problem of context specificity in regards to contextual factors (patient, physician, encounter) using well-established educational theories. Further, studies addressing clinical reasoning have primarily focused on diagnostic reasoning, and relatively few studies have explored the issue of therapeutic reasoning.¹⁴ This paper explores the influence, if any, of selected contextual factors on experts' diagnostic and therapeutic clinical reasoning. We explored experts' diagnostic and therapeutic reasoning performance using situated cognition and cognitive load theories. We expected that these two theories could help explain differences in clinical reasoning performance within videotape cases portraying the same diagnoses commonly seen in practice (context specificity).

To achieve the aims of this study, we first sought to demonstrate the problem of context specificity by comparing clinical reasoning performance as measured by our post-encounter form (PEF). This PEF is a free-text form that asks participants to list a summary statement, provide a problem list, differential diagnosis, leading diagnosis, supporting data, and treatment plan.^{15,16} We looked at PEF performance *across* videotapes portraying different diagnoses and *within* videotapes portraying the same diagnosis. Next, we explored the potential influence of each contextual factor modified in the videotapes on diagnostic and therapeutic reasoning performance. More specifically, we used multiple contextual factors and proxy measures of cognitive load (use of semantic qualifiers, problem representation, pauses, and time to establishing the diagnosis, described below) as we sought to capture if and how these measures influence clinical reasoning. Because the multiple contextual factors inherent to various clinical encounters were expected to represent information with high-element interactivity as well as goal-specific problems, we hypothesized that the presence of one or more factors would impede clinical reasoning performance due to increased cognitive load.

Methods

To measure clinical reasoning performance, we used videotape portrayals of three common conditions that internal medicine experts (board-certified internists) see in practice. Within each of these portrayals, we modified selected contextual factors. Clinical reasoning performance was captured through a free-text PEF, which served as our main dependent variable in this study. Reliability and validity evidence to support the use of the PEF for assessing clinical reasoning has been previously established.^{15,16}

We consider videotapes to be an effective method to investigate the potential influence of contextual factors on clinical reasoning because it allowed us to provide all participants with the same “exposure” to the contextual factors under study and identical case content. Further, this study design facilitated the use of a formal think-aloud procedure, which provided a measure of participants’ mental processes during the investigation. More specifically, the think-aloud protocol allowed us to explore if the selected contextual factors and our markers of cognitive load were associated clinical reasoning performance. Moreover, the videotapes provided a means to represent cases in a more authentic way than paper cases.¹⁷

To explore the potential influence of contextual factors, situated cognition⁶ would support the notion of the need for authenticity to provide a means for participant and setting interactivity (and thus information interactivity) that is present in actual clinical practice. Thus, we also asked participants to rate the authenticity of each videotape, while we systematically varied the contextual factors portrayed in the videotapes (Figure 1).

Figure 1: Videotape Portrayals and Contextual Factors Modified*

Videotape #	Contextual factors modified	Diagnosis
1.	low English proficiency	HIV
2.	Emotional volatility	Colorectal cancer
3.	low English proficiency + emotional volatility	Diabetes mellitus
4.	diagnostic suggestion	HIV
5.	atypical presentation	Colorectal cancer
6.	diagnostic suggestion + atypical presentation	Diabetes mellitus
7.	low English proficiency+ diagnostic suggestion	HIV
8.	low English proficiency+ atypical presentation	Colorectal cancer
9.	emotional volatility + diagnostic suggestion	Diabetes mellitus

*=Time to complete the PEF was varied in a random fashion for each videotape (short = up to 5 min to complete; long=up to 20 minutes to complete)

Videotape development

We constructed a series of videotapes portraying different chief complaints of three diagnoses common to internal medicine practice: HIV, colorectal cancer screening, and diabetes mellitus. For each videotape, we varied one or more contextual factors. We used a total of nine videotapes in this study (3 different presentations of these three diagnoses). Three videotapes varied selected patient contextual factors (low English proficiency, challenging credentials of physician and both low English proficiency and challenging credentials of physician), three videotapes varied selected encounter contextual factors (incorrect diagnostic or treatment suggestion, atypical presentation of condition and both incorrect suggestion and atypical presentation), and three videotapes varied one patient and one encounter factor simultaneously, in different combinations. In each videotape, we varied one or more of these patient or encounter contextual factors to explore the potential effect that these factors may have on diagnostic and therapeutic reasoning performance.

The script for each of the videotapes was written by a group of medical education experts from a variety of internal medicine specialties. These experts met on several occasions to discuss the script for each case, make revisions, and provide correct, partially correct, and incorrect response options. Two investigators (SJD and KA) then collated these response options and further revised correct, incorrect, and partially correct responses for all post encounter form sections. The videotapes were then filmed with professional actors under the guidance of a study investigator (SJD). We used the same actor to portray different manifestations of the same diagnosis (i.e. we used three patient actors) and one physician actor for all the videotapes. An independent group of physicians who were not involved in the study then reviewed the videotapes for authentic portrayal of the condition and the contextual factors represented and revisions were made, as needed.

Post-encounter form (PEF)

After viewing each videotape (viewed 3 to 6 videotapes per session), board-certified internists (experts) were asked to complete a free-text PEF. This form asked participants to list additional history and physical examination data, construct a prioritized problem list, list a differential diagnosis, list a leading diagnosis, provide supporting data for the leading diagnosis, and list a treatment plan. PEF section scores were calculated by one investigator (SJD). Each free-text answer (each additional history item, each additional physical exam item, problem list, each differential diagnosis item, leading diagnosis, and treatment plan) was scored as correct, incorrect, or partially correct response based on a separate panel of experts (board-certified internists) responses. Participants listing responses that were not present on this scoring key were vetted by two investigators (SJD and KD). Participants could list up to 5 additional history, 5 physical exam, and 7 differential diagnosis items. Participants were also given either limited (up to 5 minutes) or unlimited (up to 20 minutes) time to complete the PEF in a random fashion. The time, in seconds, needed to complete the PEF was recorded. This PEF served as a primary outcome measurement of expert's clinical reasoning performance.

Think-aloud protocol

After completing each PEF, participants completed a formal think-aloud procedure. The think-aloud procedure was incorporated into our study design so that we could determine if think-aloud utterances consistent with the contextual factors (i.e. verbalizing the presence of one or more of these factors) and our cognitive load measures, described below, were associated with clinical reasoning performance as measured by our PEF. As such, our think-aloud measures served as independent variables that were correlated with PEF performance.

More specifically, participants re-watched each videotape while “thinking aloud” with the task of establishing a diagnosis and treatment plan for the patient portrayed in the videotape. We used a standard procedure for the think-aloud protocol. Participants received initial training in this methodology¹⁶ prior to completing the protocol. All think-aloud protocols were video and audio taped. In the think-aloud procedure, participants verbalize whatever thoughts they have during a task and are explicitly instructed to *not* analyze their thought processes. Importantly, the think-aloud procedure has been shown in several studies to reliably reflect concurrent mental processes while performing a task.¹⁸

To facilitate comparison with PEF and survey measures, think-aloud protocols were analyzed according to a deductive coding system devised by three instigators (SJD, CvdV, LS) to reflect participant’s cognitive load. Because pausing the think-aloud process to measure cognitive load with traditional assessment items⁷⁻⁹ would be expected to compromise the think-aloud procedure,¹⁸ we created a coding system to ascertain real-time cognitive load while performing the task (thinking-aloud).

Measuring cognitive load

Expertise is believed to be associated with both the “chunking” of data, which is thought to free up working memory resources and by using information stored in long-term memory.¹⁸ Both of these strategies are believed to reduce cognitive load. We expected that use of proper semantic qualifiers, deep problem representation, and early establishment of diagnosis in the think-aloud procedure would equate with lower cognitive load. On the other hand, incorrect use of semantic qualifiers, superficial problem representation, pauses, and late time to establishing the diagnosis during the think-aloud would be predicted to equate with higher cognitive load; by not invoking long-term working memory resources, these measures would be expected to involve greater information processing demands on working memory.¹⁸ Likewise, verbally recognizing the presence of a contextual factor would be associated with higher cognitive load, as verbalization could be associated with the contextual factor taking up working memory space. In a previous study, we found that verbalizing one or more contextual factors during a think-aloud procedure was associated with missing key findings, which further supports our stance that high cognitive load may be negatively associated with expert’s clinical reasoning performance.¹⁹

Therefore, the presence of the following utterances were included in our think-aloud coding scheme to infer cognitive load: semantic qualifiers (treated as continuous data; positive or correct use, negative or incorrect use), deep versus superficial problem representation (binary score), verbalizing the presence of one or more of the contextual factors (yes or

no for each factor), the presence of one or more participant pauses during the think aloud (yes or no), and time to establishing the diagnosis (coded as early on in the think-aloud or late during the think-aloud). We anticipated that coding scheme of high cognitive load would be significantly associated with lower PEF section scores and vice versa.

The presence of these markers for higher cognitive load were expected to impede clinical reasoning performance and could provide additional substantiating evidence of the impact of contextual factors on expert performance as measured by PEF section scores.

Data analysis

Descriptive statistics were performed on all measures. Correlation analysis was used to determine if context specificity was present and to explore associations between measures (survey of recent exposure to correct diagnosis, cognitive load think-aloud measures, PEF section scores). Presence of context specificity was assessed through assessing PEF section correlations within videotapes portraying the same diagnosis as well as within videotapes portraying different diagnoses. We explored associations between measures to determine if contextual factors and our cognitive load measures were associated with clinical reasoning performance as measured by our PEF. Pearson correlation was used for these analyses. Additionally, to explore PEF section performance within a given series of videotapes portraying a diagnosis and across series of videotapes portraying different diagnoses, ANCOVA was used. Time to complete the PEF was used as a covariate as we suspected that time would significantly influence results. All analysis were performed with SPSS version 18.

This protocol was approved by the Uniformed Services University Institutional Review Board, and all participants signed an informed consent document prior to participation.

Results

A total of 25 (17 males, 8 females) board-certified internists participated in this study. Mean age was 39 ± 7 . Mean authenticity of videotapes was rated high by participants: 4.7 (range 3-5) on 5.0 scale (1=not authentic, 3=neutral, 5=very authentic). Performance on each section of the PEF (additional history, additional physical exam, problem list, differential diagnosis, leading diagnosis, justification for leading diagnosis, treatment plan) varied widely, ranging from 0% to 100%. Also, section scores varied on the individual participant level for each section (range 0 to 100%). Mean scores for all participants are shown in Table 1. Think-aloud data revealed that the participants acknowledged the presence of one or more contextual factors in over 96% of the videos.

Table 1: Mean Post Encounter Form Score By Section for Each Videotape Case

VT #	PEF Section	Additional History	Additional Physical Exam	Prioritized Problem List	Prioritized Differential Diagnosis	Leading Diagnosis	Supporting Evidence	Treatment Plan	Total Score
1	Mean	6.13	2.93	.92	2.73	1.29	2.71	1.07	24.60
	Std. Dev	1.92	1.79	.289	1.62	.49	3.20	1.10	4.82
2	Mean	3.27	3.33	1.33	3.80	1.70	6.07	1.73	24.35
	Std. Dev	1.53	1.72	.52	1.94	.68	3.17	.47	1.31
3	Mean	3.40	3.67	1.88	3.33	1.82	4.20	1.95	23.06
	Std. Dev	3.02	1.88	.34	1.63	.41	3.67	.16	6.1
4	Mean	5.31	3.63	1.30	3.72	1.50	4.41	1.60	22.20
	Std. Dev	1.58	2.06	.32	2.16	.54	3.51	.52	7.42
5	Mean	5.38	3.50	1.59	5.97	1.97	8.53	1.90	29.57
	Std. Dev	2.22	2.25	.42	2.09	.13	3.37	.21	6.74
6	Mean	4.00	3.31	1.43	3.84	1.29	4.72	1.18	20.42
	Std. Dev	1.633	2.750	.495	2.392	.579	2.921	.668	8.14
7	Mean	5.19	3.94	1.18	4.25	1.07	4.44	1.43	23.67
	Std. Dev	2.04	1.65	.32	2.16	.18	2.37	.68	5.62
8	Mean	5.81	4.13	1.06	4.59	1.75	3.78	1.73	23.63
	Std. Dev	1.47	1.26	.81	1.85	.40	2.39	.48	3.72
9	Mean	6.63	4.84	1.47	6.00	1.32	2.78	1.10	26.35
	Std. Dev	2.22	2.36	.39	2.03	.61	2.37	.66	5.68
Total	Mean	5.03	3.71	1.37	4.27	1.52	4.65	1.51	24.47
	Std. Dev	2.25	2.03	.53	2.21	.54	3.39	.69	6.38

Demonstrating context specificity

We compared performance of each PEF section across all videotapes (9 total), using time as a covariate, to determine if section scores differed across cases. Time was used as a covariate because, as expected, time to complete the PEF form was significantly related to most PEF section scores with low to moderate correlations (r values): physical exam (0.35), differential diagnosis (0.34), supporting evidence (0.40), and treatment plans (0.20).

Every section on the PEF, with the exception of physical exam score ($F=1.18$, $p=0.33$) showed highly significant differences by ANCOVA, with p values of <0.001 for all remaining sections of the PEF, except total score with a p value of 0.022; F value range (2.42-5.68). We then grouped cases by diagnosis (3 videotapes for each diagnosis) and correlated PEF section scores within each diagnosis. Correlations of PEF section scores within each of the three diagnoses were: history ($r= -0.47$ to 0.59), physical exam ($r= -0.33$ to 0.65), problem

list ($r = -0.63$ to 0.47), differential diagnosis ($r = -0.45$ to 0.57), leading diagnosis ($r = -0.62$ to 0.6), supporting data ($r = -0.44$ to 0.46), treatment plan ($r = -0.11$ to 0.58). Given the relatively low correlations for PEF sections within each diagnosis, we next conducted ANCOVA to determine if section scores by diagnosis were significantly different, using time to complete the PEF as a covariate. Significant differences across every PEF section score were seen with the exception of history and total score (Table 2).

Table 2: Performance across same content domain (same diagnosis) and PEF sections (ANCOVA, using time as a covariate)

PEF Section	P Value	R Squared
Additional History	0.072	0.059
Additional Physical Exam	0.002	0.121
Problem List	0.004	0.110
Differential Diagnosis	0.001	0.131
Leading Diagnosis	0.001	0.182
Total Supporting Evidence for Leading Diagnosis	0.001	0.207
Treatment plan	0.005	0.122
Total Score*	0.074	

*=Represents score on all sections of the PEF

Contextual factors and context specificity

Significant correlations between PEF section scores and think-aloud measures of cognitive load were seen. Of note, many of these correlations were negative. Correlations were consistent with our theoretic expectations in their direction (positive or negative). Specifically, participants who acknowledged the presence of low English proficiency had significantly lower scores on PEF history ($r = -0.38$) and physical examination ($r = -0.28$). Participants acknowledging diagnostic suggestion had significantly lower PEF scores on differential diagnosis ($r = -0.30$) and leading diagnosis ($r = -0.34$). Late time in establishing the diagnosis during the think-aloud protocol was also associated with significantly lower scores on PEF differential diagnosis ($r = -0.20$), leading diagnosis ($r = -0.23$), and supporting evidence ($r = -0.20$). Think aloud deep problem representation was significantly associated with leading diagnosis score in a positive fashion ($r = 0.34$). Finally think aloud incorrect use of medical terminology (negative semantic competence) was significantly associated with lower PEF leading diagnosis ($r = -0.50$) and total score ($r = -0.61$). Notably, no significant correlations between cognitive load think-aloud scores and PEF treatment plan were found.

Significant correlations were also found between mentioning one of the studied contextual factors on think-aloud and other cognitive load measures on think-aloud. Again, correlations were consistent with our theoretic expectations in direction (positive or negative). Diagnostic suggestion was associated with late time to diagnosis ($r = 0.35$) and superficial problem representation ($r = 0.26$). Low English proficiency was associated with volatility ($r =$

0.38), and negative semantic competence ($r = 0.58$). Volatility was also associated with negative semantic competence ($r = 0.43$) and was negatively associated with total semantic competence ($r = -0.26$). Finally, deep problem representation was negatively associated with late time to diagnosis ($r = -0.40$) and the presence of pauses ($r = -0.35$). Physician self-reported estimate of exposure, by number of cases with the diagnosis seen in the last week and last month, did not significantly correlate with any PEF section or any think-aloud cognitive load measure.

Grouping the videos by patient (challenging the physician's credentials or volatility, low English proficiency) and encounter (atypical presentation, and diagnostic suggestion) factors resulted in significantly lower leading diagnosis and treatment scores for diagnostic suggestion videos (respectively, $p = 0.001$, $F = 27.38$; $p = 0.013$, $F = 63.7$; ANCOVA). PEF leading diagnosis and treatment scores were not significantly lower when each of the other three contextual factors were individually compared with all other videos.

Discussion

In this study, we sought to explore the factors that may underlie the problem of context specificity. We explicitly modified selected patient and encounter factors, consistent with situated cognition, and examined performance using a PEF as well as surrogate measures for cognitive load (CLT theory). We found several notable findings.

We demonstrated context specificity on PEF performance. PEF section scores did not highly correlate for videotapes portraying the same diagnosis. Our findings add to earlier work on context specificity.¹ In particular, for a group of experts, we found significant differences across most sections of the PEF within a given diagnosis. Moreover, we studied the potential impact of specific contextual factors that could be partly responsible for the problem of context specificity using established theoretic frameworks. The fact that experts appear to be influenced by these contextual factors is noteworthy, and we hypothesize that intermediates (residents or registrars) and novices (students) may be even more susceptible to the influence of these and other contextual factors. Some aspect of the videotapes resulted in the phenomenon of context specificity; the main difference in the videotapes that portrayed the same diagnosis was the presence of one or more of our selected contextual factors. Therefore, we believe that a source of the observed context specificity was our manipulated contextual factors.

The contextual factors used in our study did appear to impact expert performance, with small to moderate effects, in accordance with the tenets of situated cognition and cognitive load theories. This impact was evident in the univariate negative correlations when contextual factors were mentioned during the think-aloud procedure with PEF performance. Not surprisingly, the physician factor of limited time impacted PEF performance on most sections. As busy physicians using electronic medical records can be expected to complete their notes in 3-5 minutes per patient, our finding of the effect of time has relevance.

The selected contextual factors studied correlated, as predicted by CLT, with our PEF and think-aloud measures. The correlations between selected contextual factors and PEF section performance varied. For example, low English proficiency had a greater effect on history and physical exam performance than the other contextual factors. CLT would pre-

dict that increased difficulty with obtaining the facts (history and physical) from a patient who has low English proficiency would lead to increased processing demands and thus higher cognitive load. Additionally, the correlation of low English proficiency with negative semantic qualifiers use may also be consistent with CLT, as an expert's use of negative semantic qualifiers is an expected manifestation of high cognitive load. It is notable that the low English proficiency factor did not have more of an impact on diagnostic and/or therapeutic decisions. This may be due to the experts' familiarity with these common conditions in practice and/or that the task of establishing the diagnosis and treatment invoke different areas of working memory, and thus additional working memory resources could be recruited to mitigate increased cognitive load. Studies from the neuroscience literature would support this finding of multiple areas of working memory, that could assist one's ability to deal with cognitive load.²⁰⁻²² This would also be consistent with the multi-modal or modality effect in CLT.²³

Incorrect diagnostic or treatment suggestion impacted differential diagnosis and leading diagnosis scores more than other PEF sections. From a CLT standpoint, suggesting an incorrect diagnosis would be expected to provide additional information processing burden related to the task of establishing the diagnosis if the diagnosis was believed to be reasonable to consider. Notably, comments related to the patient challenging the physician's credentials was frequently cited on think-aloud, but the presence of such comments was not associated with lower PEF scores. The reasons for this are unknown, but may include being readily dismissed by participants/not seriously considered.

We believe these findings have several potential implications. First, obtaining an accurate history and physical examination is essential to arriving at the correct diagnosis and therapy. Though low English proficiency did not impact diagnosis and therapy in this study, it would be expected do so in practice given the essential nature of history and physical. Second, the impact of diagnostic and/or treatment suggestion is relevant given the increasing Internet and television resources for patients to search for potential diagnoses and treatments to then discuss with their physician. Third, other contextual factors (patient, physician, and setting) that were not studied (Appendix 1) may likewise have effects on physician reasoning, either through increased cognitive load or other means. Studies exploring the influence of other contextual factors could help further disentangle the components of the problem of context specificity. Such studies could lead to modifiable solutions to reduce cognitive load such as improving physician support measures like increased appointment time.

It is also notable that negative semantic qualifier score had large, significant negative associations with several sections of PEF performance. Our results revealed that inappropriate use of semantic qualifiers (negative semantic competence) by experts appears to have a greater impact on clinical reasoning performance than proper use of semantic qualifiers. Semantic qualifiers are believed to represent a way to organize or "chunk" knowledge, enabling the participant to free up cognitive resources to attend to other information. Further supporting this view of cognitive load was the notion that negative semantic qualifiers were not a consistent manifestation for any participant during the think-aloud protocol—it was consistently manifest only in think-aloud protocols whereby the presence of one or more contextual factors was mentioned.

It is interesting that the contextual factors explored in this study appeared to impact diagnostic more than therapeutic reasoning. Though this was not a focus of the present study, this effect may be the result of several causes. First, though often believed by many to be different manifestations of the same “skill”, diagnostic and therapeutic reasoning may actually be different. For example, recent studies of the history and physical exam, two components of the clinical encounter that were long assumed to represent one “skill”—history taking and physical examination—suggest that the two are different.²⁴ Our data would likewise support the notion that the history and physical examination are quite different as correlations of history and physical exam scores were low across all videotapes (data not shown). This notion of diagnostic and therapeutic reasoning as being separate but related entities is consistent with the Reporter Interpreter Manager Educator (RIME) evaluation framework, which considers interpreting (diagnostic reasoning) as a different developmental step than managing (treatment reasoning).²⁵

Participant’s self-reported experience with the diagnosis in the last week and last month was not correlated with performance. This contrasts with recent work of novices.²⁴ This effect of experience may be more important during expertise acquisition (medical student novices) than in established experts; the latter having robust scripts for the common diagnoses used in the study.

The fact that grouping of cases by the contextual factor that was modified only led to significant performance differences in leading diagnosis and therapy with incorrect diagnostic suggestion does not mean that the other contextual factors did not impact performance as all of our videotape cases modified at least one contextual factor. What can be said is that incorrect diagnostic suggestion appears to potentially have the greatest effect on diagnostic and therapeutic reasoning performance. When grouped by diagnosis, comparing performance on videotapes when two contextual factors were present resulted in significantly lower diagnostic and therapeutic PEF scores consistent with this notion.

Theoretical implications

Several findings support the idea that cognitive load may be a reason why contextual factors impacted expert performance. Situated cognition would predict that other factors in the encounter (Appendix 1) could induce cognitive load. First, the negative impact of contextual factors modified in the videotapes was more likely to be significant when the participant verbally acknowledged the factor during think-aloud procedure. Second, several of our other think-aloud measures for cognitive load were associated with significantly lower PEF section performance, most notably the use of negative semantic qualifiers, which showed a large effect. Third, several of our cognitive load measures were significantly associated with the verbal acknowledgment of a contextual factor. These findings would suggest that measures taken that may reduce cognitive load⁷⁻⁹ could dampen the potential negative impact of contextual factors on diagnostic and therapeutic reasoning.

Our study had several limitations. First, we included a relatively small group of inter-nists. Second, performance with videotapes may not represent actual patient care scenarios and our findings therefore may not generalize to patient care settings. Third, our findings represent association and not causation. In addition, we did not include residents or medical students and similar findings may not be found in these trainee groups. Fifth, we did not

include published measures of cognitive load, and, therefore, though we found evidence to suggest that the effect was due to cognitive load, we could not confirm this result. Future work should attempt to corroborate our think-aloud data with established cognitive load measures. And finally, we included board-certified internists in our sample; the effects may be different in different specialty groups.

In all, we think this study has helped to improve our understanding of the nature of context specificity. We have shown that this phenomenon exists with modifying specific contextual factors while controlling for case content and the diagnoses portrayed in the videotapes. Further, our findings provide evidence to tie this vexing problem to cognitive load and situated cognition theories. As such we believe our work provides important theoretic as well as practical insights for untangling the problem of context specificity.

References

1. Norman GR, Tugwell P, Feightner JW, Muzzin LG, Jacoby LL. Knowledge and clinical problem solving. *Med Educ.* 1995;19:344-56.
2. Eva KA, Neville AJ, Norman GR. Exploring the etiology of content specificity: factors influencing analogic transfer and problem solving. *Acad Med.* 1998;73:S1-5.
3. Elstein AS, Shulman LS, Sprafka SA. Medical problem solving: a ten-year retrospective study. *Eval Health Prof.* 1990;13:5-36.
4. Eva, KW. On the relationship between problem-solving skills and professional practice. In: *Elaborating Professionalism Innovation and Change in Professional Education*. London: Springer; 2011:17-34.
5. Context [definition]. In: Merriam-Webster Online Dictionary [Internet]; [cited 2010 Jan 21]. Available from: <http://www.merriam-webster.com/dictionary/context>.
6. Durning SJ, Anthony R, Artino AR, Pangaro LN, van der Vleuten C, Schuwirth L. Redefining context in the clinical encounter: implications for research and training in medical education. *Acad Med.* 2010;85:894-901.
7. van Merriënboer JJG, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ.* 2010;44:85-93.
8. Paas F, Renkl A, Sweller J. Cognitive load theory and instructional design: recent developments. *Educ Psychol.* 2003;38:1-4.
9. Plass JL, Moreno R, and Brunken R. *Cognitive Load Theory*. New York (NY): Cambridge University Press; 2010.
10. Bordage G, Lemieux M. Semantic structures and diagnostic thinking of experts and novices. *Acad Med.* 1991;65:S70-2.
11. Bordage G. Elaborated knowledge: a key to successful diagnostic thinking. *Acad Med.* 1994;69:883-5.
12. Lindstom A, Villing J, Larsson S, Seward A, Aberg N, and Holtelius C. The effect of cognitive load on disfluencies during in-vehicle spoken dialogue. In: *Proceedings of Interspeech*. 2008.
13. Villing J. Dialogue behavior under high cognitive load. In: *Proceedings of SIGDIAL 2009: the 10th Annual Meeting of the Special Interest Group in Discourse and Dialogue*. 2009;322-325.
14. Monajemi A, Rikers RMJP, Schmidt HG. Clinical case processing: a diagnostic versus a management focus. *Med Educ.* 2007;41:1166-1172.
15. Durning SJ, Artino AR, Boulet J, van der Vleuten CPM, LaRochelle J, Arze B, Schuwirth L. Making use of contrasting participant views of the same encounter. *Med Educ.* 2010;44:953-61.
16. Durning SJ, Artino AR, Boulet J, LaRochelle J, van der Vleuten CPM, Schuwirth L. The feasibility, reliability, and validity of a post-encounter form (PEF) for evaluating clinical reasoning. *Med Teach*. Accepted 2011.
17. La Rochelle J, Durning SJ, Pangaro LN, Artino AR, Boulet J, van der Vleuten CPM, Schuwirth L. The effect of increasing authenticity of instructional format on student performance: a prospective randomized trial. *Med Educ*. Accepted 2011.
18. Ericsson KA, Charness N, Feltovich P, Hoffman RR, eds. *The Cambridge Handbook of Expertise and Expert Performance*. New York (NY): Cambridge University Press; 2006.
19. Durning SJ, Artino AR, van der Vleuten CPM, Pangaro L, Schuwirth L. Context and clinical reasoning: understanding the situation from the perspective of the expert's voice. *Med Educ*. Accepted 2011.
20. Anderson ML. Embodied cognition: a field guide. *Artif Intell.* 2003;149:91-130.
21. Baddeley AD. *Working Memory*. Oxford: Clarendon Press; 1986.
22. Baddeley AD. *Working memory*. *Science*. 1992;256-9.
23. van Merriënboer JJG, Sweller J. Cognitive load theory and complex learning: recent developments and future directions. *Educ Psychol Rev.* 2005;17:147-77.
24. Clauser BE, Balog K, Harik P, Mee J, Kahramen N. A multivariate generalizability analysis of history-taking and physical examination scores from the USMLE step 2 clinical skills examination. *Acad Med.* 2009;84:S86-9.
25. Pangaro L. A new vocabulary and other innovations for improving descriptive in-training evaluations. *Acad Med.* 1999;74:1203-7.
26. Durning SJ, LaRochelle J, Pangaro LN, Artino AR, Boulet J, van der Vleuten CPM, Schuwirth L. Does the authenticity of preclinical teaching format affect clinical clerkship outcomes: a prospective randomized cross-over trial. *Teach Learn Med*. In review 2011.

Chapter 7

Context and Clinical Reasoning: Understanding the Situation from the Perspective of the Expert's Voice

Accepted for publication in:
Medical Education, 2011

Steven J Durning, Anthony R Artino Jr., Louis N Pangaro, Cees van der Vleuten, Lambert
Schuwirth

Abstract

Objectives: Prior work has found that a physician's clinical reasoning performance varies on a case-by-case (situation) basis; this is often referred to as context specificity. To explore the influence of context on diagnostic and therapeutic clinical reasoning we constructed a series of videotapes to which physicians would respond, modifying different contextual factors (patient, physician, and setting). We explored how these contextual factors, as displayed by videotape encounters, may have influenced the clinical reasoning of board-certified internists (experts). Our purpose was to clarify the influence of context on reasoning, to build upon educational theory, and to generate implications for educational practice.

Methods: Qualitative data about experts were gathered from two sources: think-aloud protocols reflecting concurrent thought processes while board-certified internists viewed videotape encounters, and free-text queries that explicitly asked these experts to comment on the influence of selected contextual factors on their clinical reasoning processes. These data sources provided both actual performance data (think-aloud responses) and opinions on reflection (free-text answers) regarding the influence of context on reasoning. Results for each data source were analyzed for emergent themes and then combined into a unified theoretic model.

Results: Several themes emerged from our data and were broadly classified as components influencing the impact of contextual factors, mechanisms for addressing contextual factors, and consequences of contextual factors on patient care. Themes from both data sources had good overlap, indicating that experts are somewhat cognizant of the potential influences of context on their reasoning processes; the notable exceptions were the themes of missed key findings, balancing goals, and the influence of encounter setting, which appeared in the think-aloud analysis but not in the free-text queries.

Conclusions: Our unified model is consistent with tenets of cognitive load, situated cognition and ecological psychology theories. A number of potentially modifiable influences on clinical reasoning were identified.

Introduction

Clinical reasoning involves both establishing the diagnosis (diagnostic reasoning) and deciding on a plan for action that is specific to the patient's circumstances and preferences (therapeutic reasoning). Reasoning is generally considered a critical component of nearly everything physicians do in practice.^{1,2} Clinical reasoning enables physicians to take "wise" action; that is, to select the best-judged action in a specific situation or *context*.^{3,4}

The complex nature of clinical reasoning has been previously described.² Reasoning involves judgment, insight, and experience, all of which depend, in part, on tacit knowledge. Tacit knowledge is neither visible nor tangible, and it is difficult to evaluate.⁵ Additionally, studies have found it challenging to disentangle knowledge, experience, and reasoning.⁶⁻⁸ Further, physician performance across situations (or cases) has been shown to vary—this is often referred to as the problem of context specificity.^{1,7} Indeed, many scholars have argued that clinical reasoning is revealed only in action,⁹⁻¹⁰ and such action necessarily incorporates the potential influence of the unique context of the clinical encounter.

Context specificity poses challenges to our understanding of clinical reasoning in several ways. First, reasoning in the "real-world" (versus a controlled experimental setting) can be highly complex and dependent upon the specific patient's problems and setting in which the patient is evaluated. Work outside of medicine has found that reasoning in such authentic situations places significant reliance on the environment.^{11,12} In other words, this work suggests that the environment (or context) plays a key role in the reasoning process. Therefore, exploring the influence of context could help illuminate our current understanding of clinical reasoning, build upon existing educational theory, and provide implications for practice. What is more, recent work has shown that different participant groups can have similar, yet slightly different, interpretations of the same clinical event.¹³ This finding also suggests that context is critically important to clinical reasoning.

The influence of context on clinical reasoning is an area that has not been rigorously explored in prior studies. In short, we believe our understanding has not gone much beyond identifying context specificity. Indeed, this problem has been labeled as "the one truth in medical education".¹⁴ Furthermore, most scholars have failed to employ a common theoretical framework through a series of systematic studies. As a result, we have yet to systematically capture, unravel, or codify this complex phenomenon.

Two contemporary psychological theories, situated cognition and ecological psychology, propose a framework for considering context in both education and the clinical encounter.^{12,15-17} These theories argue that knowledge, reasoning, and learning are situated, or *located*, in experience. Experience refers to not only the participants but also the physical environment or setting. Situated cognition argues that there are complex interactions between the individual participants and the outcome (in this case, patient care) in the medical encounter. Dividing components of the clinical encounter into physician, patient, and practice factors, as supported by situated cognition, can help illuminate the potential complex interactions between these factors (Appendix 1). Adding to the situated cognition view, ecological psychology proposes agent (participant)-environment interaction in terms of affordances (what the environment provides) and effectivities (what the participant can potentially do in that environment). Affordances and effectivities cannot be completely separated (as they are interdependent) and are predicated on the specific situation. For

example, a car is drivable for an adult because the adult has the necessary effectivities, but a car is not drivable for a young child who lacks these effectivities. Finally, these two theories propose that the environment is a part of the signal, or meaningful variance to be measured, as opposed to “noise” to be minimized and generally ignored. In other words, one cannot fully appreciate clinical reasoning without considering the context or situation in which it occurs, since reasoning is located in the details of the experience at hand (participants, setting, and their interactions). In sum, these theories provide a framework to explore the context specificity problem in clinical reasoning (Appendix 1).

In this study, we explored the verbalized views of board-certified internists (experts) in relation to how selected contextual factors influenced their clinical reasoning. We did this to better understand how these contextual factors may impact their clinical reasoning and to compare what experts said to what they actually did. To do so, we analyzed both think-aloud data and free-text answers about how contextual factors may have influenced experts’ clinical reasoning; think-aloud data has been found to provide reliable and valid information about participants’ concurrent thought processes.⁶ We expected that themes emerging from experts’ after-the-fact responses “in reflection” to explicit questions (free-text answers) would largely align with themes from experts responses “in action” during the actual task (think-aloud). Further, we believed both sources of data would improve our understanding of the influence of context on clinical reasoning and that these two data sources could build on existing educational theory, providing important implications for practice.

Methods

Participants

Board-certified internists from our institution were invited to take part in the study. These participants included general internists, as well as subspecialists in internal medicine. Informed consent was obtained prior to participation, and the study was approved by the Institutional Review Board at the Uniformed Services University.

Design

To untangle the influence of context on clinical reasoning, we used a qualitative analytic approach. More specifically, we constructed a series of videotapes (of approximately 3-5 minutes) using actors from our standardized patient panel depicting three common conditions seen in internal medicine practice (colorectal cancer, HIV positivity, and diabetes mellitus). Board-certified internists (experts) viewed these videotapes without access to any other clinical information (e.g., a written medical record) about the patient portrayed. We modified selected patient and encounter factors (Table 1) to explore the influence of context on experts’ diagnostic and therapeutic reasoning. We specifically varied two patient factors: (1) emotional volatility (manifest as challenging the physician’s credentials) and (2) level of English language proficiency. We also explicitly explored two practice factors: (1) the patient’s suggesting an incorrect diagnosis and/or therapy (incorrect suggestion) and (2) typical versus atypical presentation of the condition.

In each videotape, one or both of the patient or practice factors were modified. The patient care practice setting of the videotapes was also varied, with patient care presentations in either the clinic or the Emergency Department. In each study session, three videotapes were viewed (Table 1). Successive videotapes altered one of the two factors and then both factors (patient or practice). In other words, for patient factor videotapes, video 1 depicted a patient challenging the physician's credentials, video 2 depicted low English proficiency of the patient, and video 3 depicted both challenging the physician's credentials and low English proficiency (Table 1).

Table 1: Videotape Portrayals and Contextual Factors Modified

Video #	Contextual factors modified	Diagnosis	Location*
1.	low English proficiency	HIV	C
2.	emotional volatility	colorectal cancer	E
3.	low English proficiency + emotional volatility	diabetes mellitus	C
4.	incorrect suggestion	HIV	C
5.	atypical presentation	colorectal cancer	E
6.	incorrect suggestion+ atypical presentation	diabetes mellitus	C
7.	low English proficiency+ incorrect suggestion	HIV	C
8.	low English proficiency+ atypical presentation	colorectal cancer	E
9.	emotional volatility + incorrect suggestion	diabetes mellitus	C

VT=videotape; emotional volatility=challenging the physician's credentials; incorrect suggestion= incorrect diagnostic and/or treatment suggestion

*=C (outpatient clinic); E (Emergency Department)

Participants watched videotapes and then completed a previously validated post-encounter form by writing down their diagnosis and management plan.^{13,18} Immediately after completing the post-encounter form, participants reviewed the same videotape encounter using a think-aloud protocol conducted by a research assistant.⁶ After completing the post-encounter forms and think-aloud protocols for all three videotapes, participants then completed a survey (open-ended questions). In the survey, participants were asked to reflect on the influence of the selected patient and encounter factors modified in the videotapes. However, the participants were *not* told that these factors were modified in the study. These free-text queries explicitly asked participants about the potential influence of patient and encounter factors on their practice. The specific questions were:

1. What are your thoughts about whether factor x (e.g., a patient with low English proficiency) influences your diagnostic decisions? Why or why not?
2. What are your thoughts about whether factor x (e.g., a patient who suggests an incorrect diagnosis) influences your therapeutic decisions? Why or why not?

Data analysis

The qualitative data from the open-ended survey questions were analyzed in conjunction with the think-aloud data. We used a constant comparative approach,¹⁹ whereby three investigators (SJD, TA and LS) conducted iterative coding of utterances and then grouped these utterances into themes. Following initial coding, the three investigators met to discuss themes and resolve differences. Several common themes emerged, and the process continued until saturation was reached. The themes from our analysis subsequently underwent member checking to verify the coding structure and its inclusiveness. The three investigators met for a final time to review themes and resolve differences.

Results

A total of 25 (17 males, 8 females) board-certified internists participated in this study. Mean age was 39 (SD = 7). All participants completed a think-aloud protocol for each videotape and answered the free-text, contextual factor influence questions after all the videotapes had been viewed. Saturation was reached during the analysis of both think-aloud and free-text transcripts.

From the experts' perspective, the four contextual factors investigated (two patient and two encounter factors) resulted in overlapping, yet somewhat distinct themes from the two data sources. These themes could be broadly classified into 3 categories: *components* influencing the impact of contextual factors, *mechanisms* for addressing contextual factors, and *consequences* of contextual factors on patient care.

Components influencing the impact of contextual factors

Themes grouped into this category were the requirement for additional resources, dismissing or using the contextual factor, and the influence of encounter setting. Within each theme, we also identified several subthemes.

The requirement for additional resources theme pertained to how the presence of our four contextual factors under study (low English proficiency, volatility, incorrect suggestion, and atypical presentation) often led the participant to verbalize the need for additional time, consultation, or additional testing to establish the diagnosis or provide therapy. Importantly, participants verbalized either germane or more peripheral issues that led to their use of additional resources, and this theme was present in both think-aloud and free-text analyses. Germane issues were those that were directly related to establishing the diagnosis or treatment:

'This one is tough, and I am not sure so my final diagnosis would rest heavily on additional work-up (think-aloud, atypical); 'The patient's language barrier will make me ask more questions and summarize findings more because I am less confident that I have accurate data' (free-text, low English proficiency)

Peripheral issues were those not deemed essential for arriving at the diagnosis or therapy but that still led to additional resource use. Volatility and incorrect suggestion often resulted in a peripheral issue that led to increased resource use.

'You have to take a lot more time in the appointment to address someone who questions your expertise which is distracting from the main issue at hand which is why the patient is there in the first place' (think-aloud, volatility); 'though it may be unlikely, since the patient brought it up, I often would pursue the diagnosis with additional laboratory or radiographic work-up because they expect it' (free-text, incorrect suggestion)

Notably, all four contextual factors studied could result in either germane or peripheral issues leading to increased resource use:

Since I don't speak the patient's language, I am concerned that they may not trust me (think-aloud, low English proficiency)

Dismissing or using the contextual factor was a theme in which experts also voiced several sub-themes related to how they deal with the contextual factor. These sub-themes were: credibility of the source of the contextual factors, emotions invoked by the contextual factors, and goals for the appointment. This theme and these sub-themes were present in both think-aloud and free-text analyses. Notably, participants had different opinions regarding all of these sub-themes. For example, participants did not fully agree on credibility of the source of the incorrect diagnostic suggestion:

'When the patient's diagnostic or therapeutic ideas are from a valid external authority such as WebMD or physician family member or friend, I am more likely to consider the information provided' (free-text analysis)

'I agree with the other physician's advice—yeah that's a good idea' (think-aloud)

'I felt like rolling my eyes when (the patient) said Oprah; I don't reference Oprah for my diagnostic considerations' (think-aloud)

Using one of the four contextual factors could be beneficial to the encounter or potentially problematic:

'It is helpful to explore why the patient challenges your credentials as it can provide important insights into their diagnostic or therapeutic needs' (think-aloud, volatility, beneficial use)

'I get upset'; 'I want to give up' (free-text, volatility, problematic use)

Dismissing or disregarding the contextual factor was also observed on both analyses:

'I play the odds, common things are common despite what the patient suggests' (free-text, incorrect suggestion)

Participants verbalized different goals for the appointment in both think-aloud and free-text analyses. Verbalized goals fell into two categories being pragmatic or having an orientation toward learning.

'I focus on what needs to be done to best serve the patient, given my time constraints as there are other patients to see as well' (think-aloud, atypical presentation)

'I often see the patient's suggestion as a prompt to read more so that I can readdress the suggestion at the next appointment either in person or by telephone ... ultimately this should improve my practice' (free-text, incorrect suggestion)

The influence of encounter setting was only evident on think-aloud analysis. Particularly in acute care locations, this theme appeared to represent an important component influencing the impact of our four contextual factors on clinical reasoning. The following representative comments were made regarding the influence of volatility. Like the other three contextual factors, it appears that encounter setting can dampen the participant's perception of volatility or magnify it.

'It gets more of my attention because the patient is in the emergency room (ER)—it must be more concerning to the patient'; 'Given that the patient's in the ER, and may be more concerned about her symptoms, I'd give her benefit of doubt with expressed negative emotions about me or others'; 'the patients symptoms are already more concerning to me because she is in the ER as opposed to the clinic'

'The patient has a lot of physician trust issues which I think are overboard for the ambulatory setting—makes me wonder what's really going on' (think-aloud, volatility).

Mechanisms for addressing contextual factors

Participants voiced two themes that pertain to general mechanisms for addressing these four contextual factors. One was an explicit approach (balancing goals between the patient and physician), and one appeared to be at times implicit (interaction of factors).

Our data suggested that the interplay of physician and patient goals may, in some circumstances, favorably affect the consideration of a patient's suggestion. For example, the idea of considering the work-up for an unlikely diagnosis is perhaps a means of establishing a therapeutic alliance (a goal) with the patient. This theme of balancing physician and patient goals was explicitly mentioned by some participants:

'The diagnosis is not likely but she is worried about it ... sometimes obtaining the inexpensive test to say, no it's not going on, is worth it as the therapeutic relationship that follows can help you say no to ordering tests or pursuing work-ups in the future that may be more costly as well as risky'.

'With these (contextual factors), their influence on my thinking depends ... I need to balance my goals and their goals and I often find myself meeting somewhere in the middle'

Finally, this theme of balancing goals also appears to evolve or be fluid:

'I don't know how I will respond to a patient's ideas or requests until I know what their goals for the appointment are and I will adjust my goals accordingly ... these goals are unpredictable and often seem to change as the appointment goes on'

The theme of interaction among factors was a second general mechanism for addressing contextual factors.

There were multiple examples from both the free-text items as well as the think-aloud data demonstrating how contextual factors' *interactions* represent a general mechanism for addressing our four contextual factors in regards to diagnostic and therapeutic reasoning. Sometimes this mechanism was explicitly recognized and sometimes the influence of this mechanism appeared to be under-appreciated by the participant:

'This can influence my reasoning and this is worse if (the patient's) problem is not in my area of expertise' (free-text, volatility); 'diagnostic suggestion can influence my decisions, especially with atypical presentations, which are unfortunately quite common' (diagnostic suggestion, free-text, incorrect suggestion)

'So the patient presents in non-classic fashion and wants 10,000 things done in a short appointment time while challenging your credentials ... it's hard to get everything done and focus when people have all these competing demands—important things can, and do, get confused when this happens.' (free-text analysis)

The above example incorporates both atypical presentation (volatility) and a physician factor (short appointment time).

Consequences of contextual factors on patient care

Two themes emerged from the data—missed key findings and calibration. The missed key findings theme was particularly likely to be present when participants acknowledged the presence of two or more of these four contextual factors on think-aloud analysis. Importantly, the missed data did not pertain to trivial findings; instead, the data consistently represent key diagnostic and/or therapeutic information, which the participant acknowledged when watching the video (the second time) for the think-aloud analysis. Examples of utterances from this theme included:

'I missed the tachycardia, tachypnea, and hypoxia the first time'; 'I missed the sudden onset and worsening of symptoms, which is key to clinching the diagnosis'; 'Didn't catch the head injury or hypoxemia the first time'; 'Missed important information regarding the family history of cancer'; 'forgot the high pulse rate'; 'I missed that the first time I

watched the video—she’s tachypnic, tachycardic and hypoxemic’; ‘I forgot about her past surgical history which does impact the differential and potential plans for work up’; ‘wow, she profoundly hypoxic and I didn’t catch that the first time around’

This data suggests that recognizing multiple contextual factors can influence the attention of the expert internist. This was also supported by the finding that participants found the contextual factor more difficult when the factor didn’t present in a way that meets their expectation, which may invoke more effortful processing.

‘I find these (contextual factors) particularly difficult to deal with when the patient doesn’t fit the usual pattern for the factor or my expectation ... for example, a very educated patient with low English proficiency’ (free text comment)

‘It can be particularly hard to process, I mean it really takes me back, when their incorrect (diagnostic or treatment) suggestion or their challenging of my credentials when the patient is a lawyer, doctor, or general ... this is not what I would expect from these groups’ (free text comment)

The theme of calibration referred to one’s opinions of the impact of the contextual factor on patient care. The perceived impact was voiced as either being beneficial to care or being a barrier or problematic to patient care. Importantly, when more than one contextual factor was voiced, neither factor was viewed as being beneficial. This occurred even when a specific participant mentioned one of those factors as being potentially beneficial when it was the only contextual factor present. Thus, the presence of multiple factors was consistently viewed by expert internists as problematic to effective patient care.

Participants were less likely to incorporate the patient’s incorrect treatment suggestions than diagnostic suggestions. This may be based on the assumption that the risk to the patient of an unneeded work-up is often less than inappropriate treatment. More importantly, participants verbalized that they were “less likely to very unlikely” to pursue a suggested incorrect diagnosis, and experts did not change their primary diagnostic consideration when the patient suggested an incorrect diagnosis.

‘I’d like to explore that diagnostic idea further though it’s low on my differential’; ‘it’s less likely, but reasonable, so I would pursue this’; ‘it’s not very likely but I’d pursue that diagnosis but I would not initiate that treatment at this time’

A consistent finding, which is in agreement with cognitive load theory’s (CLT) notion of limitations in our capacity to process information in working memory, was the idea that when more than one factor was present, the themes were viewed as a barrier to patient care (as opposed to being helpful). For example, several comments suggested that the presence of more than one of these four factors was particularly problematic and often would trigger additional work up.

‘I can have great difficulty putting all the diagnostic pieces together when a patient suggests a diagnosis before I have a chance to gather my thoughts ... if they also don’t speak

English very well this just escalates the problem for me ... I am more likely to order a consult to help put things together.'

Further, the suggestion of a diagnosis changes a goal-free problem into a goal-specific problem. CLT would predict a significant increase in cognitive load with changing to a goal-specific problem.

Several findings emerged that we believe support the generalizability of our above model. First, our specific contextual factors were consistently verbalized as being common or prevalent in internal medicine practice. From the *think-aloud* analysis, example utterances included the following:

'This isn't uncommon'; 'common in my practice'; 'common in medicine'; 'common issue for me'

From the free-text analysis:

'Often happens'; 'commonly happens'; 'I've been in this situation many times before'; 'I've seen this before'; 'happens all the time'

A second finding reflected the *difficulty*, at times, with dealing with these contextual issues:

'These cases represent a bad clinic day, which does happen to me more than I'd like'; 'I really struggle at times with dealing with these common issues' (free-text analysis)

A third finding supporting our model was the negative general influence the four contextual factors can have on clinical reasoning. For all four factors, utterances were made on both think-aloud analysis and free-text analysis regarding this general theme. Example utterances include:

'There's not going to be a translator, so I'm a little worried up front that this could be a suboptimal encounter because, due to the language barrier, I can't be assured that we really understand each other' (think-aloud, low English proficiency)

'This can negatively influence my decision-making' (free-text, volatility)

'Can lead to trap that leads clinician down the wrong path'; 'can lead to diagnostic dilemmas'; 'can lead to premature closure' (free-text, incorrect suggestion)

Discussion

There were four major findings from the analysis of our experts' voice. First, contextual factors can *interact* and this interaction is an important mechanism for how contextual factors influence clinical reasoning. This interaction can result in either the consequence of missed key data or the consequence of calibration or polarity (perception of the factor as

beneficial or a barrier to care). Further, our data suggests that calibration can and does change based on the presence of more than one factor. This notion of interaction between factors (two participants or a participant and the environment) is consistent with core tenets of situated cognition and ecological psychology theories (Appendix 1). For example, when participants identified more than one factor on think-aloud or free-text analysis, these factors were consistently seen as problematic to patient care, never benefits. However, and importantly, this was true only when the participant voiced the presence of multiple factors in the think-aloud; on videotapes portraying more than one factor (but for which the participant acknowledged only a single factor), the verbalized factor could be seen as beneficial to patient care.

A potential interpretation of these findings, coupled with the missed data theme, is informed by our understanding of cognitive load. It may be that factors are not viewed as positive, when multiple factors are known to coexist, due to the increased cognitive load induced by these factors. This was explicitly expressed by a participant in the free-text comments:

'It is really hard for me to process multiple (contextual) factors when I see patients' (free-text comment)

This situation may first lead to a negative perception (while the participant may be aware of cognitive load), which, if not addressed, may subsequently lead to missing key data in the encounter (once cognitive resources are overloaded). We believe this interpretation has important implications for patient care and training, as an expected extension would be that a trainee or non-expert (resident) is likely to be even *more* susceptible to this phenomenon (due to less-developed schemas in long-term memory). Furthermore, other factors from the situated cognition perspective, such as a physician factors (fatigue, well-being, or sleepiness) and/or practice factors (setting or familiarity with systems of care; Appendix 1) would likely increase cognitive load and could lead to the consequence of missed data. Future work should seek to confirm or refute this phenomenon by studying additional factors as predicted by situated cognition (Appendix 1). These studies could include investigations of participants from other specialties, as well as explorations of other contextual factors relating to the patient and/or setting. Also, helping physicians recognize when working memory resources are being stressed in the clinical situation and when more time and/or cognitive effort may be required. Such recognition may help reduce or prevent the negative consequences of this contextual factor effect. The potential benefits of “slowing down” decision making “when you should” have been discussed elsewhere.²⁰

Second, our results suggest that contextual factors influence clinical reasoning performance, and experts seem to be aware of this influence, unless there are too many contextual factors verbalized in the clinical encounter. On the one hand, it is reassuring to find that physicians appear to be aware of the extent to which many of these factors may influence their clinical reasoning. On the other hand, it is concerning that participants did not recognize that the presence of one or more factors could lead to the consequence of missing key presenting information for the diagnosis and/or therapy. This finding is similar to work from the self-assessment field, which has shown that physicians who are not competent also tend to be less aware of their difficulties.^{21,22} In other words, our data would suggest that

physicians may “miss” the available performance data regarding their own difficulties (a potential reason for being less aware of their own difficulties) due to high cognitive load. Further, this issue of missed data was seen with experts in our study. An important implication of this finding is that missed data may be even more prevalent among trainees (fellows, residents, and medical students); we believe this implication should be specifically explored in future research. Supervisors and educators have always been mindful of the potential effect of missed data on clinical reasoning. We have found that experts can articulate, in retrospect, their awareness of missing data and of confounding contextual factors. Therefore, we would hypothesize that a trainee’s explicit awareness of one or more of these factors during their presentations may be a manifestation of developing expertise. If this hypothesis is true, then through the use of additional probing questions the trainee’s clinical supervisor could improve the chances that key information will *not* be missed. Moreover, by using additional probing questions, the trainee’s clinical supervisor could find reassuring evidence that the trainee is becoming a reflective expert.

Third, though each of the four manipulated factors resulted in similar themes, our data also suggest that the “pattern” of the themes by factor can be different. For example, credibility of the source of the factor was a theme that was unique to incorrect suggestion and volatility. Incorrect suggestion and volatility were more likely to be dismissed as opposed to used by the expert. As we identified times when experts were influenced by incorrect suggestion, an implication is the need to explore the influence of diagnostic suggestion on trainees. We would hypothesize that the tendency to change one’s leading diagnosis(es) based on an incorrect suggestion occurs more frequently with less-expert participants and/or with scenarios that are uncommon in clinical practice.

Fourth, experts’ views of a given contextual factor appear complex. For example, the individual participant’s view of calibration (beneficial/positive versus problematic/negative factor for patient care), dismissing or using, germane or peripheral issue, and the influence of setting varied. A potential implication is that these views of the contextual factor may be modifiable by faculty development efforts. More specifically, addressing improved awareness of the potential impact of contextual factors, providing/teaching effective strategies for dealing with these factors, and reinforcing the importance of self-regulation throughout a physician’s practice could ultimately help to improve clinical reasoning. For example, in terms of improving awareness, not all experts recognize when a contextual factor is present. Indeed, an expert’s *recognition* as well as *reaction* to a particular contextual factor appears to be highly dependent upon other factors unique to the particular situation, corroborating the notion that context plays a complex and significant role in day-to-day patient care:

‘How does this (contextual factor) influence my decision making ... well it all depends upon what else is going on in this particular patient’s encounter; my response would depend upon everything else—did I get enough sleep, how busy is my clinic, how much time do I have, where is the appointment taking place, what other problems does the patient need to have addressed’ (free-text response)

Our data would suggest that initiatives addressing physician factors that may hinder awareness, such as sleep deprivation and burnout, could not only improve physician processing of contextual factors but also could improve physician recognition of contextual factors.

Additionally, the voiced influence of the encounter setting has important implications. For example, our findings would suggest that for a given diagnosis, physicians could underestimate the seriousness in clinic and overestimate its seriousness in the Emergency Department; this theme could represent a source of unwanted variance in medical care. Educating physicians on the potential influence this phenomenon of contextual factors on clinical reasoning could mitigate its effect.

Toward a theory of context and clinical reasoning

We believe our findings from the experts' voice add to our understanding of the influence of contextual factors on diagnostic and therapeutic clinical reasoning. Indeed, the rich and diverse comments and viewpoints regarding the impact of the contextual factors, as well as identifying multiple *components* influencing the impact of contextual factors, multiple *mechanisms* for addressing contextual factors, and multiple *consequences* of contextual factors on patient care support the presence of context specificity in clinical reasoning. Additionally, the consequences of calibration and missed key findings in the face of multiple verbalized contextual factors (when two or more challenges are acknowledged by the participant) add to our understanding of this phenomenon. Further, the mechanism of "balancing goals" of physician and patient provides additional insights into underpinnings of the vexing problem of context specificity. In all, these findings suggest that expert's view of these factors is very specific to the participants in the encounter, their balancing goals, and the situation. These insights are consistent with theories of situated cognition and ecological psychology, as depicted in Appendix 1, and with a recent definition of context.²³

Importantly, we believe our findings, which are consistent with these theories, enlighten our understanding of potential mechanisms underpinning clinical reasoning context specificity. Ecological psychology provides a rationale when the physician is aware of the influence of contextual factors—the notion of how people balance multiple, often competing goals and how these goals can change based on affordances and effectivities perceived by the clinician in the environment, which includes interactions with other participants in the environment. Further, these affordances and effectivities would be anticipated to differ based on the individual physician and additional specifics of the situation (Appendix 1). Balancing goals that often change during the encounter may be a prominent reason for why contextual factors can influence diagnostic and therapeutic reasoning in such diverse ways, as represented by our themes. Cognitive load theory²⁴ provides a rationale for what happens when the physician is unaware of the influence of contextual factors (through missing key information and at times when the mechanism of interaction is present). Cognitive load is also consistent with the notion of germane and peripheral issues that led to the requirement for resources theme. Cognitive load theory proposes three forms of load; two of these (germane and extraneous) are consistent with our findings—germane and peripheral issues, respectively.²⁴ We believe that studies exploring additional theoretic *components*, *mechanisms*, and *consequences* that may underpin context specificity represent important areas for future investigation. A number of each of these categories may be present, given the number of interactions suggested by situated cognition (Appendix 1).

Our complicated model proposes how context may influence clinical reasoning. Indeed, our model provides a rationale for how a given chief complaint in a specific setting

with specific participants can lead to diverse clinical reasoning diagnostic and therapeutic decisions. We would suggest that just like a chess master, participants demonstrating expert performance are able to quickly assess the chess board and identify important patterns, making better first moves than non experts;⁶ or, in the case of a clinical encounter, they are better able to recognize and appropriately attend to the nesting of specific physician, patient, and encounter factors and their interactions to arrive at reasonable diagnostic and therapeutic plans.

There were several limitations to our study. We sampled only a small number of inter-nists, and though saturation was reached, our findings may not generalize to other specialties in medical practice. Additionally, the use of videotapes as a trigger for participant responses may lead to different findings than in actual patient care. We also acknowledge that the themes identified are abstractions, which may indicate our own perspectives rather than actual phenomena.²⁵ The fact that consensus was reached between coders exploring two sources of data for each participant (free text and think-aloud), and the successful use of member checking, would dampen the influence of these biases, unless they are shared. We also limited our study to experts in internal medicine, and more specifically those certified by the American Board of Internal Medicine, and so our findings may be less applicable to trainees or medical students. These limitations represent important areas for future investigation.

There were several strengths to our study. In nearly every think-aloud, participants made specific utterances regarding the factors being manipulated in the study, even though they did not know the contextual factors that were being explored. Additionally, saturation was reached, and agreement between coders (and upon member checking) was good, both of which add to the trustworthiness of our findings. We also believe that the consistency of our findings with established educational theory is an additional strength.

In summary, our findings suggest that expert physicians see clinical reasoning as situated in a specific situation. Our results also suggest that specific patient and encounter factors can, and do, influence the physician's clinical reasoning process. Further, our data suggest that the influence of factors on any given physician is not uniform, and the presence of multiple factors can lead to complex interactions on clinical reasoning. Additionally, we found evidence to support tenets of situated cognition, ecological psychology, and cognitive load theories as applied to the clinical encounter, and we have proposed a model based on how context may influence clinical reasoning. Finally, our data suggest that sometimes the physician is aware of this influence (themes seen with both think-aloud and free-text) and sometimes they are not (missed key data, calibration, presence of multiple interactions and influence of setting), and that competing goals may play an important role in how these factors impact expert's clinical reasoning.

References

1. Eva KA, Neville AJ, Norman GR. Exploring the etiology of content specificity: factors influencing analogic transfer and problem solving. *Acad Med*. 1998;73 Suppl:S1-5.
2. Higgs J, Jones MA, Loftus S, Christensen N, eds. *Clinical Reasoning in the Health Professions*. 3rd ed. Oxford: Butterworth Heinemann; 2008.
3. Cervero RM. *Effective Continuing Education for Professionals*. San Francisco: Jossey-Bass; 1988.
4. Harris IB. New expectations for professional competence. In: Curry L, Wergin JF and Associates, eds. *Educating Professionals: Responding to New Expectations for Competence and Accountability*. San Francisco: Jossey-Bass; 1993:17-52.
5. Feltovich PJ, Barrows HS. Issues of generality in medical problem solving. In: Schmidt HG, De Volder ML, eds. *Tutorials in Problem-based Learning: A New Direction in Teaching the Health Professions*. Van Gorcum: Assen; 1984:128-42.
6. Ericsson KA, Charness N, Feltovich P, Hoffman RR, eds. *The Cambridge Handbook of Expertise and Expert Performance*. New York (NY): Cambridge University Press; 2006.
7. Eva K. On the generality of specificity. *Med Educ*. 2003;37:587-8.
8. Higgs J, Jones M, eds. *Clinical Reasoning in the Health Professions*. 2nd Ed. Oxford: Butterworth-Heinemann; 2000:3-14.
9. Charlin B, Tardif J, Boshuizen HPA. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad Med*. 2000;75:182-90.
10. Brailovsky C, Charlin B, Beausoleil S, Coté S, van der Vleuten CPM. Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency: an experimental study on the script concordance test. *Med Educ*. 2001;35:430-6.
11. Chaiklin S, Lave J. *Understanding Practice: Perspectives on Activity and Context*. New York (NY): Cambridge University Press; 1993.
12. Lave L, Wenger E. *Situated Learning: Legitimate Peripheral Participation*. New York (NY): Cambridge University Press; 1991.
13. Durning SJ, Artino AR, Boulet J, van der Vleuten CPM, La Rochelle J, Arze B, et al. Making use of contrasting participant views of the same encounter. *Med Educ*. 2010;44:953-61.
14. Eva KW. On the relationship of problem solving skills and professional practice. In: Kanes C, ed. *Elaborating Professionalism: Studies in Practice and Theory*. New York (NY): Springer; 2010.
15. Wilson BG, Myers KM. Situated cognition in theoretical and practical context. In: Jonassen D, Land S, eds. *Theoretical Foundations of Learning Environments*. Mahwah (NJ): Erlbaum; 1999.
16. Young MF. Instructional design for situated learning. *Educ Technol Res Dev*. 1993;41:43-58.
17. Bredo E. Reconstructing educational psychology: situated cognition and Deweyian pragmatism. *Educ Psychol*. 1994;29:23-35.
18. Durning SJ, Artino AR, Boulet J, La Rochelle J, van der Vleuten CPM, Schuwirth L. The feasibility, reliability, and validity of a post-encounter form (PEF) for evaluating clinical reasoning. *Med Teach*. Accepted 2011.
19. Kennedy TJJ, Lingard LA. Making sense of grounded theory in medical education. *Med Educ*. 2006;40:101-8.
20. Moulton CA, Regehr G, Mylopoulos M, MacRae HM. Slowing down when you should: a new model of expert judgment. *Acad Med*. 2007;82 Suppl:S109-16.
21. Eva KW, Regehr G. Self-assessment in the health professions: a re-formulation and research agenda. *Acad Med*. 2005;10 Suppl:S46-54.
22. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*. 2006;288:1057-60.
23. Durning SJ, Anthony R, Artino AR, Pangaro LN, van der Vleuten CPM, Schuwirth L. Redefining context in the clinical encounter: implications for research and training in medical education. *Acad Med*. 2010;85:894-901.
24. van Merriënboer JJG, Sweller J. Cognitive load theory and complex learning: recent developments and future directions. *Educ Psychol Rev*. 2005;17:147-77.
25. Greasley K, Ashworth P. The phenomenology of 'approaching to studying': the university student's studies within the life world. *Br Educ Res J*. 2007;33:819-43.

Chapter 8

The Effect of Increasing Authenticity of Instructional Format on Student Performance: A Prospective Randomized Trial

Accepted for publication in:
Medical Education, 2011

Jeffrey La Rochelle J, Steven J Durning, Louis N Pangaro, Anthony R Artino Jr.,
Cees van der Vleuten, Lambert Schuwirth

Abstract

Purpose: To investigate the relationship between authenticity of instructional formats and outcome measures within a pre-clerkship clinical reasoning course.

Methods: We conducted a randomized, prospective, crossover study with second-year medical students taking a pre-clerkship clinical reasoning course. Students were randomized to small groups, and exposed to three formats of differing instructional authenticity (paper case, DVD presentation, and standardized patient (SP) presentation) across three subject areas (abdominal pain, anemia, and polyuria). Three student cohorts were taught using one instructional format per subject area with each cohort receiving a different instructional format for each of the three subject areas. Outcome measures (objective structured clinical exam (OSCE), video quiz, and written exam) were selected to determine the effect of each instructional format on student's clinical reasoning performance.

Results: Increasingly authentic instructional formats did not significantly improve clinical reasoning performance across all outcome measures and subject areas. In contrast, the percent of total points from the video quiz did achieve significant differences in the anemia subject area comparing paper and SP formats (47.4 versus 57.6, respectively, $p=0.01$) and the abdominal pain subject area where the DVD format was superior to both the paper and SP formats (41.6, 34.9, and 31.2, respectively, $p=0.002$).

Conclusions: Increasing the authenticity of instructional formats does not appear to significantly improve clinical reasoning performance in a pre-clerkship course. Medical educators should balance increasing authenticity with factors like cognitive load, subject area, and learner experience when designing new instructional formats.

Introduction

Clinical reasoning refers to the synthesis of history, physical examination, laboratory and other ancillary findings into a diagnosis and plan, and represents the “bread and butter” of what a physician does.¹ Despite the essential role of clinical reasoning in the practice of medicine many questions remain regarding how expertise in this area develops.² For example, how does the instructional format impact the development of expertise in clinical reasoning? In recent decades, work from fields outside of medicine has demonstrated the importance of authentic instructional environments not only for learning, but more importantly, for *transfer* of that learning from one context to another.^{3,4} Further, transfer is believed to be closely related to the context within which material is originally learned; that is, one’s ability to apply new knowledge is strongly tied to the degree to which the new situation is similar to (or different from) the original learning context.³

Clinical education at the clerkship level typically involves instructional formats with the highest authenticity as learners see and evaluate real patients. However, pre-clerkship clinical education typically uses instructional formats of lower authenticity through a variety of surrogate patient encounters in small group settings.⁵⁻⁷ These surrogate patient encounters include reading paper cases, viewing videotaped clinical encounters between a standardized patient (SP) and physician, and viewing live clinical encounters between a SP and physician, all being followed by group discussion of the case with an attending preceptor.⁵⁻⁷

However, arguments can be made for using these less authentic surrogate encounters in the pre-clerkship years. For example, these encounters may better foster learning by simplifying the cases into key components, thus reducing irrelevant information, and aiding learners’ ability to comprehend the relevant concept.⁸ Cognitive load theory addresses the burden of information processing on learning, and refers to the ease with which information can be processed in working memory.⁸ For a given learner, each educational activity possesses an inherent degree of difficulty, or intrinsic load, which may only be modified by breaking the activity into sub-sections.⁸ Because working memory capacity is limited, too much information can inhibit knowledge from being successfully encoded into long-term memory.⁸ For example, paper cases often focus on key findings with little additional information, which may optimize cognitive load; whereas interviewing and examining an actual patient requires one to gather much more information than just the key findings, to include nonverbal and well as verbal information, which increases cognitive load. Further, the learner must attend to time pressure and perhaps the fear of asking impertinent questions with an actual patient, which can also increase cognitive load. According to cognitive load theory, unnecessary cognitive load, or extraneous load, can hamper learning because the processing of information in working memory, which is a requisite for learning, is limited; unnecessary elements can quickly overwhelm the learner.⁸ This may be especially important for the novice student who has not learned to identify key findings from additional clinical information.

While cognitive load theorists have focused on cognitive aspects of thinking and learning, others have suggested that emotional engagement is equally, if not more, important for learning.⁹ Learning environments that focus on the use of authentic learning activities assume that such activities positively impact affective factors, like motivation and emotional engagement, and thereby positively impact learning. Emotional engagement theorists posit

that engagement is key to learning—the more engaged the learner, the more learning can (and will) occur.^{9,10} In short, more authentic environments are thought to help students integrate knowledge with the attitudes and behaviors necessary for effective performance. These environments may also enhance student motivation by providing an authentic context for learning and, ultimately, enable students to transfer what is learned to novel problems encountered elsewhere.¹⁰ Additionally, highly authentic environments may assist with transfer through providing additional cues to help the learner apply a successful solution in one context to a different context. For these reasons, highly authentic environments are believed to improve learning from the perspective of emotional engagement—they enhance student motivation for learning which can increase attention to material being learned and encourage cognitive effort, ultimately leading to more effective information processing (and improved learning). However, for the novice learner, the high intrinsic load inherent to an authentic clinical encounter may yield an unacceptable degree of overall cognitive load, and thus, learning may suffer.¹¹ When considered from this perspective, it is plausible that students in their pre-clerkship training may not be ready for more authentic educational activities, and may benefit more from discrete instructional formats of lower authenticity.

Surrogate patient encounters have the potential to limit cognitive load while still providing sufficient authenticity to maximize emotional engagement. Well-designed surrogate clinical encounters for the pre-clerkship student could represent the optimal level of authenticity balancing emotional engagement, and cognitive load. There has been an ongoing call for more controlled trials and standardization of content delivered to determine the factors that actually foster improved learning with different instructional formats.^{12,13} Thus, there is a compelling need to demonstrate the impact of increasingly authentic instructional formats on student outcome measures of performance, such as clinical reasoning skills. Such data would be valuable to educators and administrators alike. Given limited resources, it is incumbent upon educators to determine if the increased cost of providing more authentic encounters may result in improved learning for trainees, or if these limited resources should be used for other trainee needs.

If the degree of authenticity in the student-patient encounter is an independent predictor of educational outcome, then one might expect to find a quantifiable relationship between increasingly authentic clinical educational formats and ensuing measures of student performance. Such a correlation might be expected to occur in a manner analogous to the classic pharmacologic “dose-response curve” where the “dose” is the level of authenticity and the “response” is the learning outcome. The purpose of our study was to investigate the relationship between authenticity of instructional formats and outcome measures within a pre-clerkship clinical reasoning course. We hypothesized that there would be a dose-response relationship between increasing authenticity and clinical reasoning performance, as measured by several outcome measurements.

Methods

Study setting

The Uniformed Services University of the Health Sciences (USU) is the United States' only federal medical school. The school has two years of preclinical education, which includes a year-long course, Introduction to Clinical Reasoning (ICR), followed by two years of clerkship education. The ICR course exposes students to a series of common symptoms, laboratory findings, and syndromes in medicine. Students are encouraged to read the material in advance of mandatory small group sessions on each topic, using whatever resources they deem appropriate. The ICR course is a blend between didactic and problem based learning (PBL) curriculum; ICR uses a case-based reasoning approach.

Within the normal context of the ICR course (<http://www.usuhs.mil/icr/index.shtml>), we selected three ICR small-group topics to test our hypothesis: abdominal pain, anemia, and polyuria. The standard educational format for the ICR course is to present subject content using paper cases, followed by small group discussion led by a faculty facilitator. Specific learning goals were derived by inter-departmental leadership for each of the cases. In terms of course timing, anemia was discussed in the fall, abdominal pain in the winter, and polyuria during the spring of the academic year; outcome measurements were all obtained in the spring of the academic year.

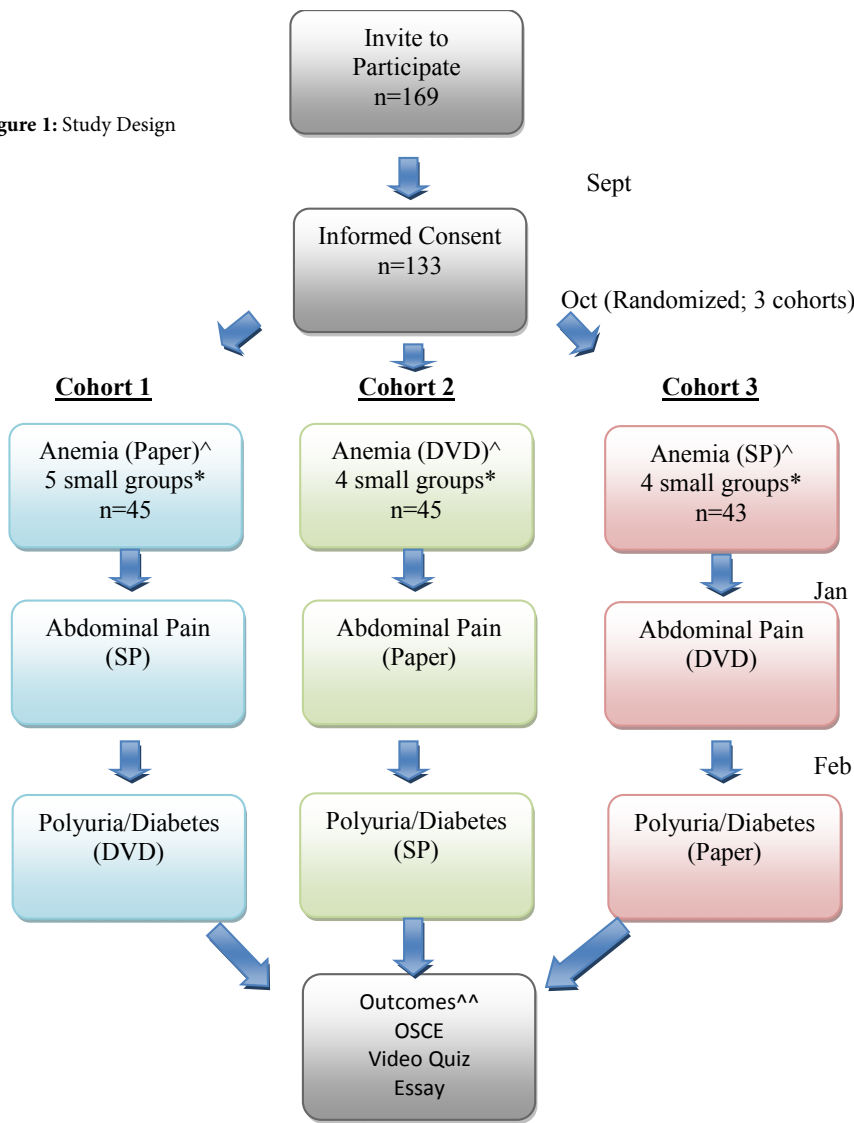
We conducted a randomized, prospective, crossover study to determine the effect of educational format on second-year medical students' performance in ICR. Students enrolled in the ICR course at USU were eligible to participate, and study subjects were randomly assigned to one of 13 small groups using a random number generator. The course directors for ICR, in addition to the individual small-group preceptors, were blinded regarding student participation in this study.

All second-year students were invited to participate. Complete participation ($n=169$) would have yielded comparison groups of 56 students for each of the three formats, which if powered at 0.80 would be able to detect an effect size of 0.53.

Educational formats

For the present study, we chose three instructional formats to present the three selected subject areas/clinical problems: paper case, DVD presentation of a doctor and patient conveying the same content, and a live SP recreation of the case. Within each of the three subject areas (anemia, abdominal pain, and polyuria), there were three cases presented per subject area over the 90-minute small group session (i.e., each case had a 30-minute discussion). Each small group was exposed to identical *content* for each clinical subject area, but was assigned to a different *instructional format* for the three cases presented within a subject area (paper, DVD, or SP; see Figure 1). Once assigned to a group, students were not able to switch between groups. In this way, study participants in each group received instruction for each subject area in only one of the three possible instructional formats, but were serially exposed to all subject areas and all instructional formats.

Figure 1: Study Design



^= Each cohort of students began with the Anemia subject area, and received teaching according to the instructional format indicated above. That cohort of students then proceeded to the remaining subject areas within this study throughout the course of the academic year. During each teaching session, students were exposed to a total of 3 cases in either paper, DVD, or SP format.

*=There were a total of 13 small groups during the study period. These 13 small groups were grouped into 3 cohorts.

^^= Each outcome measure contained three unique study cases

Within each subject area of the study, we selected existing cases from the ICR course, and further developed each study case with input from content experts in each topic area. Once fully vetted, each paper case was used as the framework to develop a script portraying the salient features of each case. To ensure consistency of case content across the instructional formats, all students received a copy of the paper case regardless of the instructional format they would experience, and the SPs on the DVD encounters used the exact same script as the SPs in the live encounters. For the DVD scenarios, students watched a doctor interview and examine a patient followed by small-group discussion facilitated by a faculty member. For the SP scenarios, students watched a faculty member interview and examine the standardized patient followed by small-group discussion facilitated by the same faculty member. The students in all three educational formats (paper, DVD, SP), for each case, had open-ended questions to complete, which shaped the subsequent discussion (i.e., What further information would you like to obtain by history and physical examination? What is your differential diagnosis?) To help ensure consistency in teaching, all small group facilitators were oriented to the study, course goals, and small group session expectations, but not to study outcome measurements. Group facilitators were provided identical discussion points to cover after the presentation of each case within the three topics, regardless of the instructional format used for that case.

Measurements

Baseline demographic and academic data (grade point average [GPA], which we used as a proxy of pre-study clinical reasoning ability) were collected on all study participants, and effective randomization throughout all the small groups was achieved using a random number generator. We collected both graded outcome measures within the normal context of the ICR course, in addition to a non-graded outcome that was added for this study. The three outcome measures were selected to match each learning format with a test format, and each outcome measure contained three unique study cases so that all students were exposed to all nine study cases without repetition between outcome measures. Details of the three primary outcomes measures are provided below, and all measures assessed medical knowledge and the objective structured clinical exam (OSCE) measured knowledge and physical exam skills; in other words, students were not assessed on procedures (e.g. lumbar puncture) on our outcome measurements. Consistent with transfer theory^{3,4} we matched outcome measurements to the teaching formats (i.e. students were assessed on paper, DVD, and SP formats).

1. An end-of-year objective structured clinical examination (OSCE) served as a graded assessment of integrated clinical skills, and matched the SP instructional format. All students were required to complete a six station OSCE as part of the ICM-III (a course during the second year that assesses students' ability to conduct a history and physical examination on a patient presenting with common medical problems) and ICR courses. Three of the OSCE stations were identical to one case from each of the three study subject areas. Students were evaluated at each OSCE station by a SP, who completed a checklist of historical and physical exam components the students were expected to correctly accomplish.

2. An end-of-year video quiz was added to the ICR course as a non-graded evaluation of clinical reasoning skills to match the DVD instructional format. Students viewed three DVD cases, one identical case from each of the three study subject areas, and were asked to provide short answer responses with appropriate justifications to include: generate a prioritized problem list, create a differential diagnosis, and provide a series of additional information they would pursue in the history, physical exam, or ancillary data. This outcome was included to attenuate the potential “noise” introduced from additional test preparation and/or study time associated with graded evaluations.
3. An end-of-year ICR written exam served as another graded assessment designed to match the paper case instructional format. This exam presented students with a series of nine paper cases from three booklets (three cases per booklet), with one booklet containing only paper cases identical to study cases (one from each subject area). Students were required to select one case from each booklet. Students were given an opportunity to read the paper case, and were then required to provide short answer responses identical to those outlined above in the video quiz above.

For each study topic (anemia, abdominal pain, polyuria), participants received one of the teaching formats (paper, DVD, SP), and crossover was established so that each participant received all three teaching formats. Figure 1 outlines the study design. Students were divided into 13 groups, and the same 13 teachers facilitated the groups across the three study topics. The small-group instructors were also responsible for evaluating student performance at the end of the session.

Data Analysis

Since each outcome had a different total score, reported scores on all outcome measures were expressed as the percentage of total points, and displayed as mean score with a 95% confidence interval. Since study outcome measures could be influenced by student’s prior academic performance, primary analysis of outcome measures was accomplished by first comparing GPA terciles across groups. We then conducted a series of two-way ANOVAs using educational format and GPA as the independent variables, and our three outcomes as dependent variables. GPA was used as an independent variable as students with higher GPAs would be expected to perform better, everything else being equal.¹⁴ Since the teaching method groups had differing numbers of students, post-hoc analysis was accomplished using Tukey HSD as a more conservative method to determine the statistical significance of any individual educational format amongst the numerous other comparisons. Lastly, as an additional analysis, all outcome measures were standardized using a Z-transformation (mean of 0 and standard deviation of 1) to facilitate comparison of teaching methods across outcome measures. ANOVA of study group demographic data and outcome measures was used to determine the effectiveness of randomization. Student’s T-test was utilized for comparisons of less than three different groups, and Chi-Square test was utilized for proportional data. All statistical analysis was accomplished using SPSS version 12. This study was approved by the USU institutional review board, and all students who participated signed an informed consent.

Results

Demographic Data

Of the 169 second-year medical students enrolled in the course, 133 (78.6%) consented to participate in this study. There were a total of 13 small groups with an average of 13 students in each group. Baseline demographic and prior academic performance measures for each cohort were collected by a research assistant, who de-identified all study participant data (see Table 1).

Table 1: Baseline Demographics and Non-Study OSCE Stations Scores*

	Baseline Demographics			Non-Study Related OSCE Stations		
	Number of Students	1 st Yr GPA	Gender Distribution	Oral Presentation Station	Stroke Station	Geriatrics Station
		Mean \pm SD	% Males	Mean \pm SD	Mean \pm SD	Mean \pm SD
Cohort 1	45	2.90 \pm 0.49	62%	72 \pm 16	78 \pm 10	79 \pm 7
Cohort 2	45	2.97 \pm 0.48	81%	68 \pm 15	75 \pm 13	77 \pm 7
Cohort 3	43	2.83 \pm 0.53	74%	67 \pm 14	74 \pm 13	77 \pm 8
Overall Class	169	2.92 \pm 0.50	73%	67 \pm 15	76 \pm 12	78 \pm 7

*=Data presented as mean values and standard deviations for GPA and OSCE Station scores. The three OSCE stations represent the non-study subject areas (Oral Presentation, Stroke, and Geriatrics) on the standard ICM-III OSCE, and represent a percentage of the total points available for that station. ANOVA analysis demonstrated no significant differences between the three cohorts and the overall class values for GPA or OSCE station scores. Chi-Square analysis demonstrated no significant differences in the gender distributions between cohorts or overall class gender distribution.

An average of 76% of students in each ICR small group were enrolled in the study, with an average 1st year medical school GPA of 2.92. Analysis by ANOVA demonstrated no significant differences by group on overall 1st year GPA, our pre-implementation measure of clinical reasoning ability, or by scores on non-study OSCE stations, indicating successful randomization of students across all study groups (see Table 1). First year GPA was used to determine successful randomization as this baseline variable was believed to potentially have a significant effect on our outcome measurements. Chi-square analysis demonstrated no differences across cohorts by gender further supporting successful randomization. Additionally, analysis by ANOVA demonstrated no differences across small groups with respect to any of the outcome measures, indicating that no student received either an advantage or disadvantage from being placed in a particular small group (data not shown). Furthermore, successful randomization, as demonstrated above, would be expected to correct for differences in pre-implementation clinical reasoning ability.

Student verbal and written feedback supported our authenticity model (paper was least engaging and SP was most engaging format). Verbal feedback was collected from students

following each of the small group sessions from the course directors. Additionally, the course directors collected comments regarding the small group sessions through weekly correspondence with the student representative for the course. Students are also asked following each small group session “How did the small group session go?” “What were its strengths and potential areas for improvement”? Furthermore, written feedback was obtained from end-of-trimester evaluation form data (i.e., three times per academic year) which contained free text comments asking students to make additional comments about any of the lecture or small group sessions.

Students rated the DVD and SP cases as the most engaging formats and recommended that these two formats be used exclusively with teaching the course in subsequent years. Consistent with cognitive load theory, both students and instructors commented that although the SP and DVD cases were more engaging, they were also potentially more difficult as they included additional nonverbal information which was not conveyed in the paper cases.

Outcome Measurements

OSCE Station

None of the comparisons were statistically significant (see Table 2). However, as would be expected, participants in the top third of the class by 1st year GPA outperformed their fellow students on almost every station (data not shown).

Table 2: Study Outcome Measures[^]

	Major Outcome Measures								
	OSCE Station Scores (% available points)								
	Diabetes Station			Abdominal Pain Station			Anemia Station		
	N	Mean	95% CI	N	Mean	95% CI	N	Mean	95% CI
Paper	54	72.1	69.0-75.2	43	76.8	74.2-79.4	34	57.2	53.9-60.6
DVD	34	71.6	67.3-75.9	54	76.0	73.2-78.8	43	60.0	56.7-63.3
SP	43	74.7	71.0-78.4	34	78.7	75.6-81.8	54	58.6	55.8-61.4

	Written Exam (% available points)								
	Diabetes			Abdominal Pain			Anemia		
	N	Mean	95% CI	N	Mean	95% CI	N	Mean	95% CI
Paper	15	65.5	58.2-72.8	9	70.0	62.7-77.3	14	65.4	56.9-73.8
DVD	14	66.3	61.2-71.3	21	76.2	72.5-79.9	16	68.4	62.2-74.2
SP	17	70.0	65.4-74.6	6	76.3	67.7-84.8	17	64.1	60.4-68.2

	Video Quiz (% available points)								
	Polyuria (Diabetes)			Abdominal Pain			Anemia		
	N	Mean	95% CI	N	Mean	95% CI	N	Mean	95% CI
Paper	49	39.2	34.3-43.3	35	34.9	30.8-39.2	29	47.4	39.9-54.8
DVD	29	32.4	27.0-37.8	49	41.6*	37.8-45.2	35	49.4	43.1-56.7
SP	35	36.6	31.6-41.4	29	31.2**	26.5-36.0	49	57.6†	50.9-63.6

[^]=Data presented as the means and 95% confidence interval for each outcome measure by instructional format across all three study subject areas. Since each outcome measure has a different amount of total points available, means represent the percentage of total available points for that specific outcome measure.

*p=0.053 between Paper and DVD instructional formats

**p=0.002 between DVD and SP instructional formats

†p=0.01 between Paper and SP instructional formats

ICR Written Exam

None of the comparisons were statistically significant. Comparing most authentic (SP) with least authentic (paper) instructional format, effect sizes were: 0.73 (abdominal pain), 0.09 (anemia) and 0.43 (polyuria). Comparing DVD with paper instructional formats, effect sizes were: 0.72 (abdominal pain), 0.27 (anemia) and 0.08 (polyuria). There were improved scores between the bottom and top third of students, but this was not statistically significant (data not shown).

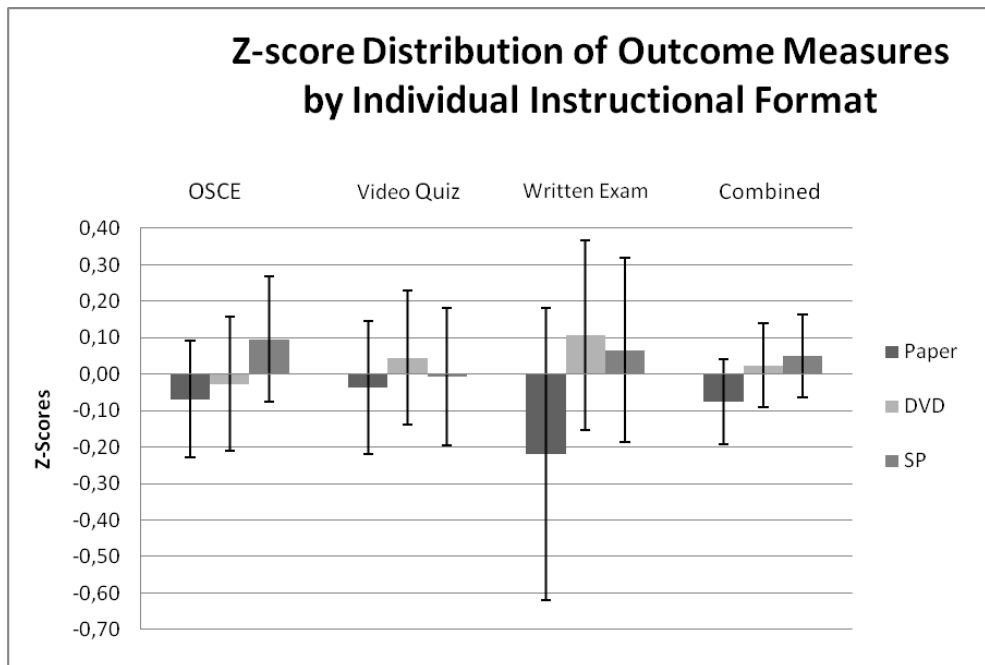
Video Quiz

In contrast to the other two outcome measures, there were statistically significant differences in the comparisons (see Table 2). Of note, ANOVA of the Abdominal Pain subject area scores demonstrated significance ($p = 0.002$) with the DVD format superior to either Paper case or SP (41.6, 34.9, and 31.2), respectively (see Table 2). Also, points from the video quiz did achieve significant differences in the anemia subject area comparing paper and SP formats (47.4 versus 57.6, respectively, $p=0.01$).

Z-Transformation

To better assess the impact of the instructional formats across all three study topics and across all three outcome measures, each case within each outcome measure was standardized using a Z-Transformation. Results of combining all outcome measures are shown in Figure 2.

Figure 2: Z-Score Distribution of Outcome Measures by Individual Instructional Format*

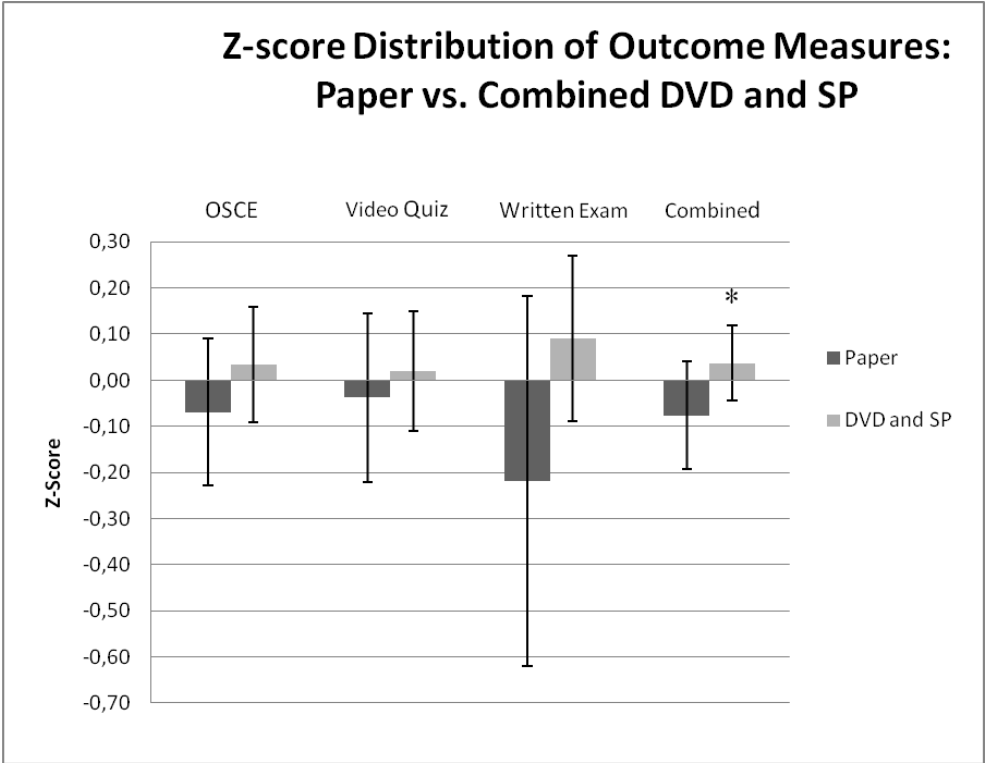


*=Data presented as the combined subject area Z-score (solid bars) and 95% confidence interval (error bars) for each instructional format across each outcome measure separately. The combined data points represent the Z-score and 95% confidence interval for each instructional format across the combined outcome measures.

Given that the paper case format typically underperformed both the DVD and SP formats, a secondary analysis of the Z-transformation was accomplished by comparing the paper case format with the combined DVD and SP formats. Again, the paper format demonstrated lower scores across all outcome measures (see Figure 3). Combining all outcome measures

demonstrated a difference in Z-scores from -0.076 to 0.037 comparing the paper case format to the combined DVD and SP formats, respectively (see Figure 3), with an effect size of 0.12.

Figure 3: Z-Score Distribution of Outcome Measures Comparing Paper with the Combined DVD and SP Instructional Format[^]



[^]=Data presented as the combined subject area Z-score (solid bars) and 95% confidence interval (error bars) for the paper compared to the combination of DVD and SP instructional formats across each outcome measure separately. The combined data points represent the Z-score and 95% confidence interval for the paper compared to the combination of DVD and SP instructional formats across the combined outcome measures.

*p=0.059

Discussion

We were unable to demonstrate a statistically significant impact of instructional format on student performance. Thus, for second-year medical students, we could not corroborate the belief that as we increase the degree of authenticity in our instructional formats, there will be a stepwise increase in learning and transfer. As seen with our study, this intuitively appealing assumption might not always hold true, and more often, the relationship between instructional format and learning appears to be more complex. The assumption that increased instructional authenticity should improve learning is consistent with theories of emotional engagement;^{3,9} however, this assumption neglects the potentially negative impacts of increasing cognitive load with authenticity.⁸ In short, the advantages that might be gained by increasing authenticity through emotional engagement could be mitigated by the disadvantages of increasing cognitive load through complex instructional formats. Our findings would suggest that educators should consider the interplay of emotional engagement and cognitive load when designing instructional formats.

This prospective randomized trial begins to explore the importance of authenticity in constructing instructional formats to foster greater emotional engagement while also addressing the potential limitations of increasing authenticity through increasing cognitive load. This study also demonstrates the complexities inherent in the methods we use to teach and assess learning.³ Our findings suggest that the instructional format used to teach second-year medical students clinical reasoning may have an impact on student performance as measured by our outcomes, though almost all of our findings were not statistically significant. Comparing the paper and SP instructional formats, the Cohen's *d* effect sizes averaged across the subject areas for the OSCE, the video quiz, and the paper quiz were 0.19, 0.31, and 0.40, respectively. Given the initial power analysis, which indicated the capability to detect an effect size of 0.53, which is a moderate effect, our study may not have had an adequate sample size. It is therefore plausible that additional participants could have resulted in statistically significant findings.

Several issues consistent with educational theory are raised by this study.^{3,8,9} First, the emotional engagement of a particular case may have an impact on whether an instructional format with increased authenticity translated into improved clinical reasoning outcome measures.^{3,9} The cases written for the three subject areas in the study had varying degrees of emotional and/or physical content based upon the actual subject area. For example, the SP abdominal pain cases had a wide range of patient discomfort from the extreme pain of pancreatitis to the mild discomfort of irritable bowel syndrome. In contrast, the anemia cases all presented with similar complaints of vague fatigue, but without any dramatic emotional content. Therefore, SP instructional formats may simply make the abdominal pain cases "come alive", and impart more emotional engagement during the teaching process than the potentially less emotionally-engaging anemia cases. In fact, the abdominal pain cases demonstrated small to moderate effect sizes with increasing authenticity of instructional format across all outcome measures, while the anemia cases demonstrated small effect sizes between the authenticity of instructional format and outcome measures with the exception of the video quiz. This finding suggests that greater emotional engagement may lead to enhanced student learning in these content areas as measured by our outcome measurements.

Second, we found evidence supporting the transfer literature regarding the importance of matching the instructional format to the transfer test, or outcome measure format.³ We used three outcome measures (OSCE, video quiz, and written exam) to match the three instructional formats (SP case, DVD case, and paper case). The effect of matching an instructional format to an evaluation format was most evident on the Video Quiz for the abdominal pain case. Students receiving the DVD educational format showed improvements over both the paper case and SP formats (See Table 2).

Third, medical students are academically successful and skilled at preparing for and taking examinations. Therefore, range restriction in exam grades makes it difficult to uncover important effects from instructional interventions. The focus on studying for graded exams could wash-out the effect of an instructional intervention on these clinical reasoning graded measures, making it difficult to determine if there was any effect from the instructional format at all. Therefore, an ungraded video quiz outcome measure was added to represent an instructional format unbiased from the impact of additional study time. We found a statistically significant increase in outcome measures within the abdominal pain subject area. However, the polyuria subject area demonstrated no differences further demonstrating the complexity of the educational process that relies on a complicated interplay between instructional format, subject area, emotional engagement, cognitive load, and transfer test format.

Given limited resources, does the increased cost of providing more authentic encounters result in improved learning for trainees? Based on our study, the answer may depend on the topic being taught and the goals of the instructional format. It may be that subject areas with low emotional content, like anemia, do not induce higher emotional engagement with increasing authenticity, and thus may be taught just as effectively using paper case discussions as compared to the more resource intensive DVD or SP formats. However, some subject areas, with higher emotional content, like abdominal pain, may induce high emotional engagement, and thus could be worth the increased cost. Further studies addressing the long-term impact of these various instructional formats may alter their value, especially if more authentic encounters in the second year of medical school continue to demonstrate positive effects into the third year clerkships.

This study had several limitations. First, we only looked at three subject areas. The clinical reasoning literature suggests content specificity is important, and would advocate for more sampling of subject areas.¹⁵ Next, we only studied students from one class year and one institution. Third, we only explored immediate outcomes at the end of the second year; third year clerkship outcomes and possibly longer term outcomes may reveal important group differences. Fourth, this study involved pre-clerkship students; findings may be different for students in their clinical years or in their post-graduate education. Specifically, studies have demonstrated that cognitive load induced by instructional format can differ dramatically based upon the learners' level of training.¹⁶ Fifth, the size of our "small groups" within the ICR course was relatively large, and our study findings could have been different if each small group had fewer trainees. Sixth, the length of time between instructional format and outcomes may have attenuated the impact of format on outcomes. Seventh, we used GPA as a pre-implementation surrogate for clinical reasoning performance. Though GPA is arguably not an optimal measure, we made the assumption that our successful randomization procedure effectively attenuated any pre-existing differences that may have

existed between subjects at the outset of the study. Eighth, though our argument for increasing cognitive load with increasing authenticity is supported by the literature, we did not directly assess cognitive load for participants, which should be pursued in a future study. Ninth, we have limited reliability and validity data on our outcome measurements. Tenth, we did not control for other student “factors” such as prior clinical experience with the disease (expertise), perception of emotional engagement, or fatigue. We also did not control for other “teacher” factors such as interest or fatigue, “patient” factors such as the complexity of the diagnosis or “encounter” factors such as learning climate. Several theories do suggest that these and other factors may contribute to the outcome and interact in dynamic ways.¹⁷⁻²⁰ We did, however, largely control for instructional method with faculty training as we sought to determine if the instructional format had influence on our outcomes. Finally, our design may have lacked power to find relatively small differences between groups. These other factors may have impacted our results.

To our knowledge, this study represents the first prospective, randomized control trial investigating how various educational formats affect clinical reasoning performance. Although there may be a positive correlation between the authenticity of an instructional format and clinical reasoning performance, this relationship appears to be complex and is likely impacted by a variety of factors to include: subject area, emotional engagement, and cognitive load. Future studies controlling for these factors may better elucidate the complex process of learning in medical school. There is an alluring attraction to the use of technology in medical education and an attempt to create more authentic instructional formats with the hope of improving overall student learning. This notion is fed by both the simulation industry and our desire to reach an ever increasingly tech-savvy student population. As our study demonstrates, increasing instructional authenticity is not a panacea. Accordingly, the use of highly authentic instructional formats should be approached with caution; clearly, more theory-driven research is needed to further elucidate its’ role in medical education.

References

1. Richardson W. Integrating Evidence into Clinical Diagnosis. In: Montori V, ed. Evidence-Based Endocrinology. Totowa (NJ): Humana Press: Mayo Foundation for Medical Education and Research; 2006.
2. Ericsson K. Deliberate practice and acquisition of expert performance: A general overview. *Acad Emerg Med*. 2008;15:988-994.
3. Bransford JD, Brown AL, Cocking RR. How People Learn: Brain, Mind, Experience, and School. Washington, D.C.: National Academy Press; 2000.
4. Norman G. Teaching basic science to optimize transfer. *Med Teach*. 2009;31:807-811.
5. Carter MB, Wesley G, Larson GM. Lecture versus standardized patient interaction in the surgical clerkship: a randomized prospective cross-over study. *Am J Surg*. 2006;191:262-267.
6. Davidson R, Duerson M, Rathe R, Pauly R, Watson RT. Using standardized patients as teachers: a concurrent controlled trial. *Acad Med*. 2001;76:840-843.
7. van Zanten M, Boulet JR, Norcini JJ, McKinley D. Using a standardised patient assessment to measure professional attributes. *Med Educ*. 2005;39:20-29.
8. Sweller J, van Merriënboer JJ. Cognitive architecture and instructional design. *Educ Psychol Rev*. 1998;10:251-296.
9. Dai DY, Sternberg RJ, eds. Motivation, Emotion, and Cognition: Integrative Perspectives on Intellectual Functioning and Development. Mahwah (NJ): Lawrence Erlbaum Associates; 2004.
10. van Merriënboer JJ, Kester L. Taking the load of a learners mind: Instructional design for complex learning. *Educ Psychol*. 2003;38:5-13.
11. Kirschner PA, Sweller J, Clark RE. Why minimal guidance during instruction does not work: An analysis of the failure of the constructivist, discovery, problem-based, experiential, and inquiry based teaching. *Educ Psychol*. 2006;41:75-86.
12. Clark R. Media will never influence learning. *Educ Tech Res Develop*. 1994;42:21-29.
13. Cook D. The research we still are not doing: an agenda for the study of computer based learning. *Acad Med*. 2005;80:541-548.
14. Gonnella JS, Erdmann JB, Hojat M. An empirical study of the predictive validity of number grades in medical school using 3 decades of longitudinal data: implications for a grading system. *Med Educ*. 2004;38:425-434.
15. van der Vleuten CP, Newble DI. How can we test clinical reasoning? *Lancet*. 1995;345:1032-1034.
16. Wetzels S. Individualized Study Strategies for Prior Knowledge Activation. The Netherlands: Open University; 2009.
17. Durning SJ, Artino A, Pangaro LN, van der Vleuten C, Schuwirth L. Redefining context in the clinical encounter: Implications for research and training in medical education. *Acad Med*. 2010;85:894-901.
18. Wilson BG and Myers KM. Situated cognition in theoretical and practical context. In: Johnassen D and Land SM, eds. Theoretical Foundations of Learning Environments. Mahwah (NJ): Erlbaum; 1999:57-88.
19. Bredo E. Reconstructing educational psychology: Situated cognition and Deweyian pragmatism. *Educ Psychol*. 1994;29:23-25.
20. Young M. Instructional design for situated learning. *Educ Technol Res Dev*. 1993;41:43-45.

Chapter 9

Does the Authenticity of Preclinical Teaching Format Affect Subsequent Clinical Clerkship Outcomes?

A Prospective Randomized Crossover Trial

Under editorial review

Steven J Durning, Jeffrey LaRochelle, Louis Pangaro, Anthony Artino Jr., John Boulet, Cees van der Vleuten, Paul Hemmer, Dodd Denton, Lambert Schuwirth

Abstract

Background: Educational theories predict conflicting results for the effect of increasing the authenticity of the teaching format of complex information on educational outcomes. We sought to determine the effect of increasingly authentic small-group, pre-clerkship teaching format on clerkship outcomes to further enlighten this debate.

Methods: Students enrolled in a prospective randomized crossover trial that involved three content areas. For each content area three teaching formats were tested. Participants were randomized to teaching format by content area. Clerkship outcomes were performance on an objective structured clinical exam, a DVD exam, internal medicine clerkship grades, and performance on the subject examination. The data were analyzed using a multivariate analysis of covariance.

Results: 133(78%) students participated. Teaching format did not have a statistically significant effect on any of the specified clerkship outcomes. However, number of patients seen was significantly associated with higher scores in respective outcomes by topic.

Conclusions: Second-year teaching format did not directly influence subsequent clerkship performance. Our study adds to the literature by demonstrating that the authenticity of pre-clinical teaching format does not appear to matter for clerkship performance; however, the number of actual patients seen does appear to influence related clerkship outcomes.

Introduction

Medical school educators grapple with many questions when selecting optimal teaching formats for students such as whether increasing authenticity of any teaching encounter (i.e., making it more akin to actual patient care) is better than alternatives such as paper cases for preclinical students. Clinical clerkship education typically involves instructional formats with the highest authenticity as learners see and evaluate real patients. Pre-clerkship education, however, typically uses instructional formats of lower authenticity through a variety of surrogate patient encounters in small group settings.¹⁻³ These surrogate patient encounters include reading paper cases, viewing videotaped clinical encounters between a standardized patient (SP) and physician, and viewing live clinical encounters between a SP and physician, all being followed by a discussion of the case with a clinical preceptor.

Grappling with teaching formats is reflected in the move toward more integrative pre-clerkship curricula. However, such curricular decisions must balance such factors as limited faculty time, human resources, and money, as well as the use of alternative, burgeoning teaching materials such as simulation, video, and virtual learning. Furthermore, the amount of material we ask medical students to learn continues to grow at a rapid pace. As such, we need empirical studies that can help us determine the best methods to educate our future physicians.

Clinical reasoning is at the heart of what a physician does--coming up with diagnostic and treatment plans that are individualized to the needs of each patient.⁴ Although we have learned much about the acquisition of clinical reasoning, many studies have not addressed the potential impact of teaching format on the acquisition of reasoning.

Three educational theories are relevant to this research question. Two of these leading educational theories would predict opposing outcomes with increasing the authenticity of teaching materials in medicine while the third could foresee either opposing outcome. Cognitive load theory deals with limitations in processing simultaneous information in working memory.⁵⁻⁸ According to cognitive load theory (and to several other mainstream cognitive psychology theories) humans can only actively process, or pay attention to, 4-7 elements of information at one time.⁵⁻⁸ Cognitive load theory also distinguishes between intrinsic and extraneous load. Intrinsic load is related to the difficulty of the task itself, the number of elements that need to be processed, and the number of possible combinations of these elements. Reduction of intrinsic cognitive load can be achieved by simplifying the task or by making use of previously stored information (automations or schemata) in long-term memory. In fact, data show that the intrinsic load associated with a certain task is largely dependent upon the existing knowledge structures of the individual learner.^{6,7} Extraneous load refers to cognitive activities that are not necessary for learning, such as searching for information that is needed to complete a complex learning task. Cognitive load theory would thus predict that increasing the clinical authenticity of a preclinical teaching encounter, by adding verbal and nonverbal information conveyed by a patient, would increase extraneous load and could lead to inferior learning outcomes in less experienced medical students as compared to simplified paper cases using key features. Further, with highly complex information (i.e., high intrinsic load), cognitive load theory would predict that a key feature approach, using paper cases, would minimize cognitive load for less experienced medical students, and thus improve learning outcomes compared with video (DVD) and

standardized patient (SP) cases where extraneous load (both verbal and nonverbal) is introduced.

On the other hand, emotional engagement theories of learning would predict that the more engaged the learner is with the material, the more likely that learner will develop meaningful understanding.⁹ In this sense, the more connected learners are to the content, the more likely they will learn. Within this theoretical framework, paper cases would be considered less authentic, leading to suboptimal learning, whereas video and live patients would lead to more optimal learning because of engagement. This idea of increasing authenticity leading to improved learning is also espoused by a third group of theories known collectively as situativity theory.¹⁰⁻¹²

Situativity theory is a more inclusive theory than either cognitive load or emotional engagement. Situativity theory argues that all learning is situated (or placed) within a context, and it takes into account the participants in an encounter and their interactions. In this framework, cognitive load (a learner factor) is interacting with emotional engagement (a patient factor as well as a patient x learner interaction); this framework also takes into account the environment (for learning or teaching). Appendix 1 illustrates the factors and their interactions proposed by situativity theory.^{13,14} Unlike emotional engagement theory, situativity theory would posit that the cognitive load of the learner could negatively affect educational outcomes and supports the importance of exploring this balance (i.e., potential interaction) between authenticity and cognitive load.

This investigation is a part of a two-year prospective, randomized, crossover study to determine the effect of different educational formats on second-year medical students' performance in a second-year preclinical education course on clinical reasoning and during their internal medicine clerkship the following year. We have previously reported findings on preclinical performance when this cohort completed their second year of medical school (the first year of the study);¹⁵ the present study involves the findings from the internal medicine clerkship.

As detailed in the previous study,¹⁵ we constructed clinical cases to use in three 90-minute sessions of an introduction to clinical reasoning (ICR) course in the second year of medical school. The three session topics were anemia, abdominal pain, and diabetes mellitus. These three topics were chosen as they are commonly seen in practice; the specific cases introduce students to the processes of clinical reasoning through a series of tasks--summary statement, problem list, differential diagnosis, leading diagnosis and a plan. For each of these three topic sessions, we created three different teaching formats (paper case, DVD with patient-doctor actor portrayals, and live standardized patient portrayal), but each case contained identical clinical information. In the prior study, we found no statistically significant differences between teaching format (paper, DVD, SP) and our main outcome measures at the end of the second year. We did find, however, that the formative DVD outcome (i.e., an outcome not counted toward final grade) yielded differences based on teaching format. We reasoned that this could be due to its ungraded (formative) nature. On this point, it is possible that students may have done better on this formative assessment because they learned the information versus simple recall of the information with cramming for a graded exam.

In the current study, we followed the same cohort through the third-year clerkships and used the clerkships as a transfer test, where transfer is defined as using material learned in

one context to solve a novel problem encountered elsewhere.¹⁶ Notably, previous studies have shown that the results of learning (transfer) can be delayed which represented a second reason that we followed this cohort longitudinally for two years.¹⁶

Prospectively, we sought to determine if second-year teaching format impacts performance on the third-year internal medicine clerkship. This study adds important clinical outcomes (transfer test findings) to help inform second-year teaching methods. We used the third-year internal medicine clerkship as a proxy for transfer of learning, and we also included outcome measures in paper, DVD, and objective structured clinical examination (OSCE) format to see if there were differences in learning based on teaching format. We hypothesized that increasing the authenticity of teaching format in a pre-clerkship clinical reasoning course would result in improved clinical clerkship performance as suggested by emotional engagement theory and seen as likely by situativity theory.

Methods

Study setting

The Uniformed Services University of the Health Sciences (USU) is the United States' only federal medical school. During the time of the study (academic years 2008-2010), our institution employed a curricular approach comprised of two years of preclinical work followed by two years of clerkship education.

The Introduction to Clinical Reasoning (ICR) course is a year-long course that exposes students to common symptoms, laboratory findings, and syndromes in medicine. This course uses a case-based reasoning approach. The standard course format involves paper-case discussions regarding the topic in small groups of 9-12 students led by a faculty member with subject matter expertise.

Three topics for the current study were used: anemia, abdominal pain, and diabetes mellitus. Although different teaching formats were used for all students for these topics, the course director and the individual small-group preceptors in the ICR course were blinded regarding those students who elected to participate in this study, as were the third-year clerkship directors and teachers in the outcomes phase of the study.

All second-year medical students were invited to participate. Only students who consented to participate in the study during the second year were included in this study of subsequent third-year outcomes. Complete participation (n=169) would have yielded comparison groups of 56 students for each of the three formats, which if powered at 0.80, would be able to detect a moderate effect size of 0.53.

Educational formats

We chose three instructional formats for study: paper case read by the student, DVD presentation of a doctor and patient conveying the same content which was watched simultaneously by all students in the small group, and a live SP presentation of the case whereby students in the group may question the SP after the portrayal but before the faculty discussion. Within each of the three conditions, there were three cases presented per condition (ane-

mia, abdominal pain, and diabetes mellitus) over the 90-minute small-group session (i.e., each case had a 30-minute discussion). Importantly, each small group was exposed to identical *content* for each condition (topic); each small group, however, was assigned to a different *instructional format* for the three cases presented within each condition. Study participants in each group received instruction for each of the three subject areas in only one of the three possible instructional formats (paper, DVD, SP), but were serially exposed to all subject areas and all instructional formats (see Figure 1; chapter 8 of this thesis).

To help ensure *consistency of content*, all students received a copy of the paper case regardless of the instructional format they would experience, and the SPs on the DVD encounters used the exact same script as the SPs in the live encounters. To help ensure *consistency in teaching*, all small group facilitators were oriented to the study, course goals, and small group session expectations, but not to study outcome measurements. Group facilitators were provided identical discussion points to cover after the presentation of each case within the three topics, regardless of the instructional format used for that case. Student participation within small-group discussions is required by preceptors, but not quantified. We did not attempt to quantify differences in student participation in small groups as a measure of engagement. We did ask about engagement of the small-group sessions on end-of-trimester evaluation forms.

Measurements

Baseline demographic and academic data were collected on all study participants. Using a random number generator, we randomized participants to each small group such that participants were evenly distributed to the teaching format intervention within a topic, and also across the three topics, to ensure each participant was exposed to each teaching format. We used randomization to mitigate against potential unmeasured differences between participants that were not collected in the study, such as prior clinical experiences, learning preferences, etc. Clerkship year outcome measures were selected to match learning formats, and included the following:

1. An end-of-third-year summative OSCE on abdominal pain. This OSCE station was identical to one abdominal pain case discussed during the second year.
2. An end-of-third-year year internal medicine clerkship, open book, summative multi-step exam.¹⁷ This is a DVD exam that is comprised of several steps. In the first step the student watches several minutes of an initial encounter between a doctor and a patient and then they are given 10 minutes to write down additional history and physical examination items that they would like to pursue. In the second step, students review the remainder of the clinical findings (whether or not they had asked for these) and then are given five minutes to construct a prioritized problem list; and in the final step students are given the same complete, standardized problem list for the patient and have twenty-five minutes to create a prioritized differential diagnosis and treatment plan. Step reliabilities range from 0.65 to 0.83. One (of three) multi-step exam cases was identical to one of the anemia cases discussed during the second year.
3. National Board of Medical Examiners (NBME) shelf exam in medicine. This is a multiple-choice examination that served as a “proxy” for a paper exam. Typical reliability for this exam is 0.75.

4. Internal medicine clerkship “clinical points”. The third-year internal medicine clerkship at USU is 12 weeks in length, with six-week ward and six-week ambulatory rotations. Anemia, abdominal pain, and diabetes are among the fifteen core problems that all students are expected to encounter during the clerkship. Teachers are guided in the evaluation process through use of a common framework (Reporter-Interpreter-Manager/Educator) and face-to-face formal evaluation sessions at all clerkship sites.¹⁷ Recommended grades are converted to points (reliability is 0.83¹⁸) and this total “clinical points” serves as a summation of all teacher evaluations of student performance over the twelve week clerkship. As anemia, abdominal pain, and diabetes mellitus are such common conditions seen in the internal medicine clinics and on the wards, this grade, in part, encompasses student performance in these domains.

As a covariate, we also included the number of patients seen during the internal medicine clerkship with each of these conditions (anemia, abdominal pain, and diabetes) from a required, student self-reported logbook. Third-year students on the internal medicine clerkship are required to keep an electronic log (Clinical Weblog) documenting the number of patients seen with 15 common conditions in internal medicine. Prior studies from our institution have demonstrated that this measure has reasonable reliability and is highly correlated with the actual number of patients seen.¹⁹

Data Analysis

As we have previously reported, first-year GPA across groups did not differ, nor did end-of-second-year outcomes not related to the current study.¹⁵ This provides important supporting data that our randomization worked. Since outcome measures could be influenced by student’s prior academic performance, our prior study also demonstrated that there was no difference in GPA terciles across groups.¹⁵ This study continued to follow this cohort through the third year.

We conducted multivariate analysis of covariance (MANCOVA) with our above outcome measurements serving as the dependent variables. To test the influence of teaching format on these outcome variables, we used the nesting of educational format (paper, DVD, or SP) and diagnosis (anemia, abdominal pain, and diabetes mellitus) as independent variables.

Numbers of patients seen as documented by the Clinical Weblog for each of the study conditions (anemia, abdominal pain, and diabetes mellitus) served as covariates. For statistically significant findings, post hoc analysis with analysis of covariance (ANCOVA) was performed. To determine size of the effect of significant variables by MANCOVA, we conducted a post-hoc linear regression of the outcome of interest to ascertain explained variance. We used SPSS version 14 for inferential statistical analysis. This study was approved by the USU Institutional Review Board, and all students who participated gave written informed consent.

Results

Demographic Data

Of the 169 eligible medical students, 133 (78.6%) consented to participate in the initial year of the study, and all of these students participated in the next year of the study reported in this paper (see Figure 1, chapter 8 of this thesis). In our prior report,¹⁵ ANOVA demonstrated no significant differences by group based on GPA indicating successful randomization.

End-of-Trimester Evaluation Forms

Eighty-eight percent of students completed these evaluation forms. Over 90% of respondents reported that the DVD and SP portrayals were more engaging and a superior learning format to the paper cases.

Outcome measurements

MANCOVA results indicated that the educational format (the primary intervention of interest) did not have a statistically significant effect on any of the outcome measurements (see Table 1).

Table 1: Multivariate Analysis of Variance

Statistic	Value	F Value	Number of degrees of freedom	Pr>F
Wilks' Lambda	0.9193	0.63	48	0.9764
Pillai's Trace	0.0816	0.63	48	0.9796
Hotelling-Lawley Trace	0.073	3.31	8	0.9718

However, each of the covariates did result in statistically significant effects on each of their respective outcomes (MANCOVA, all p values <.01). Specifically, students who reported seeing a greater number of patients during the internal medicine clerkship with abdominal pain had significantly higher OSCE scores on the abdominal pain outcome (explained 13% of variance by regression). Similarly, students who reported seeing more clerkship patients with anemia had significantly higher scores on the anemia multi-step exam case (explained 29% of variance by regression). There was no "crossover" effect seen with numbers of patients reported with one problem and its affect on performance with regard to other problems. In other words, students who reported encountering more patients with abdominal pain did not have differences in anemia or diabetes outcomes, and more patients with anemia did not impact abdominal pain or diabetes outcomes. However, reporting greater number of patients with each of these three conditions also resulted in significantly greater clinical points from teachers in the internal medicine clerkship (each variable explaining

about 2-3% variance by stepwise regression) and NBME examination performance (each variable explaining about 2-3% of variance by stepwise regression).

Discussion

Medical educators are tasked with selecting the best teaching materials to facilitate student learning, and have to decide among instructional methods of varying authenticity and resource intensiveness. Educators may use both experience and understanding of educational principles to shape their decisions. Cognitive load theory and emotional engagement predict conflicting results regarding the effect of increasing the authenticity of the teaching format on educational outcomes. We sought to determine the effect, if any, of increasing the authenticity of teaching format in a second-year clinical reasoning course on clerkship-year outcome measures. We used third-year clerkship measures as outcomes as they represent, in effect, a transfer test. Further, studies on learning have previously shown a delayed effect on some outcomes; in other words, although immediate testing may fail to show a difference in learning material, an effect may become evident as more time goes on.¹⁶ We hypothesized that increasing the authenticity of the teaching format would result in improved educational outcomes. However, we found that the authenticity of the teaching format did not result in significantly improved scores for any of our outcome measurements.

What are some potential reasons for why authenticity of material did not improve clerkship outcomes? We propose several possible explanations, which we believe would be suitable as hypotheses for future study. First, perhaps the groups were too large in the second year to enhance emotional engagement sufficiently and thus lead to improved learning. Students in groups of 9-12 may not have a sufficient “dose” of authenticity; this study could be repeated with smaller groups to address this question. Comments from students on end-of-trimester evaluation forms overwhelmingly endorsed the increased emotional engagement with DVD or SP discussions over paper case discussions; however, this effect could be much greater with smaller groups. Second, simply “watching” a DVD or intermittent questioning of an SP during the 30-minute encounter may not provide enough emotional engagement to enhance learning over paper cases; perhaps students must grapple more actively with the content to enhance their emotional engagement. The educational literature has demonstrated that active, personal engagement with content can improve learning over vicarious (observational or second-hand) experiences.^{16,20} Third, three cases with each format may not be enough exposure to the teaching format to enhance learning. For example, repeating the cases in the forms of unknowns or review cases could enhance the exposure needed to observe differences with each teaching method.

The above limitations are ameliorated by seeing patients with the condition in question on the clerkships—students are taught in a one-on-one fashion (versus a small group) with evaluating patients in the clinic or wards, and, importantly, they evaluate the patients themselves as opposed to observing someone else evaluate the patients (i.e., a vicarious experience). Clinical Weblog also revealed many more patient encounters per topic (versus only three “encounters” per topic in the course). Furthermore, actual patients represent the highest degree of authenticity, and thus our findings do suggest that increasing authenticity

can perhaps improve outcomes, though this finding should be replicated in subsequent studies.

There were several notable findings which add to the literature on this topic. To our knowledge, we conducted the first prospective, randomized, crossover study that followed participants over a two-year period to address the hypothesis about authentic teaching formats and subsequent performance. It is important to point out that our randomization was successful and consistency in content and teaching was emphasized. We did not control for other student factors such as prior clinical experience with the disease (expertise) prior to medical school, perception of emotional engagement, or well-being. However, given our successful randomization of other factors, we suspect that these factors were equivalent across groups. The key finding in this study was that the covariates (numbers of actual patients seen with our three study conditions) were significantly associated with higher performance on respective outcome measurements. When encountering real patients on clinical clerkships, students are taught in a more individualized fashion (versus a small group) and, importantly, they evaluate the patients themselves as opposed to encountering them through an observational experience.

As we followed this cohort over two years prospectively, we can say with more confidence than prior reports that exposure to patients in the clerkship year with the condition measured in the outcome does result in improved performance. This is consistent with the deliberate practice literature, which argues that to become an expert in a field, time of exposure with content, initially under the guidance of a mentor or coach, improves outcomes (performance).²¹

Additionally, our results further support the context specificity of clinical reasoning²² in that seeing more patients with abdominal pain, for example, did not improve learning outcomes for anemia or diabetes. Again, our prospective design and successful randomization adds to prior work in this field and suggests that educators need to track numbers of patients seen if they want to attempt to improve content-related outcomes during the third year.

The notion that increasing authenticity of the teaching format can improve learning is also an important tenet of situativity theory.¹⁰⁻¹² Situativity theory argues that learning, knowledge, and thinking is situated in experience. From this standpoint, the patient, the physician, and the environment can all contribute to the outcomes measured and can do so in non-linear (or more than additive) ways as the interactions between these three perspectives also impact performance. This theory also suggests additional factors (patient, physician, and environment), that were not controlled, and that could have impacted our outcomes.^{13,23} For example, the learning climate for the cases, individual student well-being, and group dynamic factors all could have impacted student learning.

Cognitive load theory would suggest that the improvement in performance on the third-year clerkships with additional exposure to the conditions could be the result of having more knowledge and experience and thus being better able to manage the additional cognitive load imposed by actual patients. Cognitive load theory would also support testing of the hypothesis that actual patient encounters in the second year improve performance. One cannot assume that less-experienced students would be able to handle even greater potential cognitive load with seeing actual patients during the second-year, despite our

findings that show the number of actual patients seen may improve performance during the third year.

There were several limitations to our study. We only included three subject areas, and we did not explicitly measure cognitive load or emotional engagement. By controlling for content in the paper cases, DVDs and SPs, we designed a study with escalating cognitive load (additional verbal and nonverbal information that was designed to reinforce and not distract from the content of the cases) though we did not measure the participant's view of this in real time. We also assumed that more authentic portrayals (i.e., paper vs. DVD or SP) would result in more emotional engagement. Our findings from student written feedback supports this notion of increasing emotional engagement with moving from paper to DVD or SP, but we did not sample this for all students during all three cases per session. We only studied students from one class year and one institution. Our study involved clerkship students; findings may be different for students in their fourth year or in their post-graduate education. Specifically, studies have demonstrated that cognitive load induced by instructional format can differ dramatically based upon the learners' level of training.⁵⁻⁸ The size of our small groups within the ICR course may have been too large to allow optimal engagement. Though the clerkship served as a transfer experience, the length of time between instructional format and outcomes (one year) may have attenuated the impact of teaching format on outcomes. We did not control for other "teacher" factors such as interest in teaching the content or "patient" factors such as the complexity of the diagnosis or "encounter" factors such as learning climate as well as the mitigating factor of student preparation for summative exams.

Notwithstanding these limitations, our study suggests that while second-year teaching format may not matter in terms of third-year outcomes, the number of patients seen during the third year-clerkship (the most authentic format) does. Additional studies should further explore the balance between cognitive load and emotional engagement and their potential implications for learning.

References

1. Carter MB, Wesley G, Larson GM. Lecture versus standardized patient interaction in the surgical clerkship: a randomized prospective cross-over study. *Am J Surg.* 2006;191:262-267.
2. Davidson R, Duerson M, Rathe R, Pauly R, Watson RT. Using standardized patients as teachers: a concurrent controlled trial. *Acad Med.* 2001;76:840-843.
3. van Zanten M, Boulet JR, Norcini JJ, McKinley D. Using a standardized patient assessment to measure professional attributes. *Med Educ.* 2005;39:20-29.
4. Richardson W. Integrating evidence into clinical diagnosis. In: Montori V, ed. *Evidence-based Endocrinology.* Totowa (NJ): Humana Press: Mayo Foundation for Medical Education and Research; 2006.
5. Sweller J, van Merriënboer JJG, Paas F. Cognitive architecture and instructional design. *Educ Psychol Rev.* 1998;10:251-96.
6. Plass JL, Moreno R, Brunken R, eds. *Cognitive Load Theory.* New York (NY): Cambridge University Press; 2010.
7. van Merriënboer JJG, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ.* 2010;44:85-93.
8. van Merriënboer JJG, Kirschner PA, Kester L. Taking the load off a learner's mind: instructional design for complex learning. *Educ Psychol.* 2005;38:5-13.
9. Dai DY, Sternberg RJ, eds. *Motivation, Emotion, and Cognition: Integrative Perspectives on Intellectual Functioning and Development.* Mahwah (NJ): Erlbaum; 2004.
10. Kirshner D, Whitson JA, eds. *Situated Cognition: Social, Semiotic, and Psychological Perspectives.* Mahwah (NJ): Erlbaum; 1997.
11. Robbins P, Aydede M, eds. *The Cambridge Handbook of Situated Cognition.* Cambridge (MA): Cambridge University Press; 2009.
12. Young MF. Instructional design for situated learning. *Educ Technol Res Dev.* 1993;41:43-58.
13. Durning SJ, Artino AR, Pangaro LN, van der Vleuten CPM, Schuwirth L. Redefining context in the clinical encounter: implications for research and training in medical education. *Acad Med.* 2010;85:894-901.
14. Durning SJ, Artino AR, Holmboe E, Beckman TJ, van der Vleuten CPM, Schuwirth L. Aging and cognitive performance: challenges and implications for physicians practicing in the 21st century. *J Contin Educ Health Prof.* 2010;30:153-160.
15. La Rochelle J, Durning SJ, Pangaro LN, Artino A, van der Vleuten CPM, Schuwirth L. The effect of increasing authenticity of instructional format on clinical reasoning performance in a pre-clerkship medical student course: a prospective randomized crossover trial. *Med Educ.* Accepted 2011.
16. Bransford JD, Brown AL, Cocking RR. *How People Learn: Brain, Mind, Experience, and School.* Washington (DC): National Academy Press; 2000.
17. Durning SJ, Pangaro L, Denton GD, Hemmer PA, Wimmer A, Grau T, Gaglione M, Moores L. Inter-site consistency as a measurement of programmatic evaluation. *Acad Med.* 2003;78:S36-8.
18. Roop SA, Pangaro L. Effect of clinical teaching on student performance during a medicine clerkship. *Am J Med.* 2001;110:205-9.
19. Denton GD, Williams RW, Pangaro L. Core problems reported by students in a palm OS and Internet-based problem entry system predicts performance on the third-year internal medicine clerkship. *AMIA AnnuSymp Proc.* 2003;827.
20. Shell DF, Brooks DW, Trainin G, Wilson KM, Kauffman DF, Herr LM, eds. *The Unified Learning Model: How Motivational, Cognitive, and Neurobiological Sciences Inform Best Teaching Practices.* New York: Springer; 2010.
21. Ericsson KA, Charness N. Expert performance. *Am Psychol.* 1994;49:725-47.
22. Eva K. On the generality of specificity. *Med Educ.* 2003;37:587-8.
23. Durning SJ, Artino AR, Boulet J, Van der Vleuten CPM, La Rochelle J, Arze B, et al. Making use of contrasting the views of participants in the same encounter: are they the same? *Med Educ.* 2010;44:953-61.

Chapter 10

Using Functional Neuroimaging Combined with a Think-aloud Protocol to Explore Clinical Reasoning Expertise in Internal Medicine

Under editorial review

Steven J Durning, John Graner, Louis N Pangaro, Anthony R Artino Jr., Thomas Beckman, Eric Holmboe, Terrance Oakes, Michael Roy, Gerard Riedy, Vincent Capaldi, Robert Walter, Cees van der Vleuten, Lambert Schuwirth

Abstract

Introduction: Clinical reasoning is an essential aspect of medical practice but is difficult to assess. Functional magnetic resonance imaging (fMRI) and think-aloud protocols offer potential means for better understanding of clinical reasoning. Therefore, we used a combination of fMRI and think-aloud procedures to determine if there were fMRI correlates to analytic and non-analytic reasoning theory and to determine if there are fMRI differences with answering versus thinking aloud on vignette-based test items.

Methods: Board certified internists (experts) answered, and reflected on, validated United States Medical Licensing Exam (USMLE) and American Board of Internal Medicine (ABIM) multiple-choice questions (easy and difficult) during a fMRI scan. This procedure was followed by completion of a formal think-aloud procedure.

Results: Seventeen internists completed the protocol. Evidence, in the form of fMRI changes, to support the presence of analytic and non-analytic reasoning systems was found. Specific fMRI findings revealed significant activation of prefrontal cortex when answering incorrectly versus correctly ($p < .01$) and activation of precuneus and mid-temporal gyrus when not guessing versus guessing ($p < .01$). Some activation differences were found when comparing answering versus reflecting upon answers.

Discussion: We found fMRI evidence to support existing analytic and non-analytic reasoning theory, and the use of think-aloud procedures to capture internal mental processes. This model of employing fMRI correlates may enhance our understanding of theory, as well as our practice of teaching and assessing clinical reasoning.

Introduction

Expertise in clinical reasoning is essential for optimal patient care.^{1,2} Over recent decades, there has been considerable debate regarding the best model for describing how experts engage in clinical reasoning.^{3,4} Clinical reasoning involves both establishing the diagnosis and determining a care plan that is tailored to the individual patient. Indeed, clinical reasoning can be thought of as the sum of the thinking and decision-making processes associated with clinical practice.² However, studies have shown that performance on one clinical case does not necessarily predict performance on other cases. This vexing finding is referred to as problem of context-specificity.⁵⁻⁸

The problem of context specificity, or the variation in performance from one situation to the next, has challenged our ability to study clinical reasoning.⁸ This challenge is due, in part, to the nature of clinical encounters, which tend to be highly complex and dependent upon the unique settings and particular patients' problems.⁹ Additionally, the investigation of clinical reasoning necessarily requires inferences about physicians' internal mental processes, using tools that were designed to measure observed behaviors, which are assumed to reflect processes within the brain. Indeed, until recently, scientists have generally avoided serious discussion and investigation of the functional activation of complex psychological constructs such as cognition.¹⁰

Recently, however, techniques have been developed that may allow us to observe complex internal mental processes, such as clinical reasoning, more directly. One such technique is functional magnetic resonance imaging (fMRI). Functional MRI is a non-invasive technique that enables the investigator to assess regions of brain activity by identifying quantifiable differences in brain perfusion. Regions of the brain that use more oxygen (to process stimuli, an "artifact" of thought) are believed to evoke a rapid feedback loop that dilates blood vessels to such areas, resulting in an increase in oxygenation. The change (relative increase or decrease) in oxygenation levels can be reliably measured by fMRI.

A second method for making inferences about internal mental processes is the think-aloud protocol. Think-aloud protocols have been used in multiple prior studies exploring expertise.¹¹ These protocols allow participants to express whatever thoughts come to mind while performing a task. In a think-aloud protocol participants are specifically instructed to *not* analyze their thoughts while performing the task of interest; instead, they are asked to engage in the task and verbalize whatever comes to mind. This type of verbal expression of thoughts does not appear to change the underlying structure of the participant's thought process.¹¹ However, the validity of the think-aloud procedure is still an issue of ongoing debate since some methodologists believe that the act of generating the report may change the very cognitive processes that mediate the observed performance.¹¹

One approach to further studying the validity of a think-aloud procedure is to couple it with fMRI analysis. Such an approach could provide insights into cognitive processes underlying clinical reasoning. Further, by combining event-related fMRI findings with a think-aloud protocol and performance measurements, we can potentially establish links between cognition and its neural substrate. Functional MRI methodology provides a potential means for testing the hypothesis that there is a functional neuroimaging basis to educational theory. To date, however, the technique has not been used for medical education purposes, nor have studies been conducted that assess complex mental processes in experts.

Therefore, to enhance our understanding of clinical reasoning, we paired fMRI with a think-aloud protocol to assess expert performance on existing gold-standard measures of clinical competence, using the lens of educational theory to assist with interpretation.

Educational theory

The literature has recently called for studies that attempt to link theory with neuroimaging findings.¹²⁻¹⁴ A recent review suggests that physicians use two primary systems for clinical reasoning.¹⁵ One system is rapid, automatic, and unconsciously controlled; it is typically referred to as non-analytic reasoning.¹⁵ The second system is slow, rule-based, and consciously controlled; it is typically referred to as analytic reasoning.¹⁵ Studies have suggested that physicians may use both systems during patient care.^{4,15}

Script theory provides one, prominent hypothesis regarding how physicians reason. Scripts are knowledge structures of specific diagnoses and treatments held in physicians' long-term memories. When a rich, highly organized, script exists in a physician's memory for a patient's presentation, the non-analytic system is activated; this process is called "pattern recognition". This fast, non-analytic system is believed to be used predominantly by experts, whose knowledge is presumably more organized in memory and embedded in rich scripts.¹⁶⁻¹⁸ In the second process (analytic reasoning) physicians consciously select diagnoses, explicitly seeking additional data to confirm or refute potential diagnoses. Analytic reasoning is believed to occur when selecting between two or more possible diagnostic options, or when the physician is unsure of the diagnosis altogether. In other words, analytic reasoning refers to the careful analysis of the relation between symptoms and signs and diagnoses. Experts (in medicine) are believed to use both analytic and non-analytic systems with predominance for non-analytic reasoning.^{11,16-18}

The non-analytic system's features of rapid retrieval and limited effortful information processing are related to the fundamental concept of working memory.¹⁹ In essence, each of our sensory inputs (sight, sound, proprioception, etc.) generates neural outputs that are sent to working memory. Working memory provides temporary storage and manipulation of this information, which is necessary for such complex cognitive tasks like reasoning.²⁰ Since there is often too much sensory input for clinicians to deal with simultaneously, a primary function of working memory is to choose what of this input will be ignored and what will be processed.¹⁹

Magnetic resonance imaging and reasoning

Until recently, fMRI had been used primarily in conjunction with simple on/off stimuli such as pictures and words; whereas more complex mental processes, such as problem solving, have been avoided due to the multifaceted responses that could be engendered in the brain.¹⁰ A roadmap is needed to guide the construction of experiments and interpretation of findings for complex mental processes. Educational theory provides a potential roadmap for achieving this. Indeed, recent calls for converging-methods approach, such as our current study design, has been suggested as the optimal way to use fMRI to explore complex mental processes.¹²

Functional MRI studies from fields outside of medicine have explored patterns of brain activation with simple and more complex tasks and have shown conflicting results including increases, decreases, and sometimes both increases and decreases, in regions of brain activation.^{11,21} Localized increases in regional brain perfusion result in increased brain activation on fMRI. There are also some areas of the brain where decreases in blood flow appears to occur, which is thought to represent inhibition of these areas, perhaps enabling more rapid and efficient responses due to a more focused, or specialized, area of activation. Studies with simple tasks (letter recognition and simple auditory tasks)²¹ have shown increased cortical tissue activity devoted to the task after periods of training (“specialized” to the task), while other studies have shown functional reorganization,¹¹ which involves reorganization of brain activity or activity of different brain areas during different stages of learning. This work would suggest the following two themes that are applicable to this study: 1) learning appears to be localized and very specialized, and 2) learning can produce both increases and decreases in areas of activation, with the belief that decreased activation reflects increased efficiency in data processing, and increased activation reflects increased cortical tissue devoted to the task.²²⁻²⁴ Furthermore, studies from fields outside of medicine suggest that working memory processing may occur in multiple specific locations in the brain.^{20,25} Importantly, decreased activation and increased efficiency of data processing could be consistent with non-analytic reasoning; increased cortical tissue devoted to a task could be consistent with analytic reasoning.

There is limited fMRI evidence from fields outside of medicine to support a dual-process system--a slow system and a fast system. The slow system is often referred to as the “control network” or “domain general network”. This system may equate with high effort (analytic) reasoning, but this interpretation remains controversial and may be task specific.^{21,26} Further, this theory has not been explored within medicine or with experts within a field. This “control network” area is believed to involve portions of the frontal, cingulate, and parietal cortices. The fast system is postulated to reside in a subset of these areas as early experience may require support of this control processing system,²⁴ which then wanes as practice (and learning) occurs. Thus, it is thought that the activity in control processing regions should decrease (or drop out) with practice.²⁴ Other studies involving simple tasks after a period of practice have implicated the prefrontal, anterior cingulate and parietal-temporal cortex²⁷ with the fast system. Additionally, the caudate and/or precuneus have also been implicated in learning and memory.²⁸

The purpose of our present study was to determine if specific fMRI findings are associated with clinical reasoning expertise as measured with vignette-based multiple-choice questions (MCQ). In other words, does analytic and non-analytic reasoning theory in medicine have a neuroimaging foundation? Does expertise in medicine have a neuroimaging foundation? More specifically, we explored potential differences in fMRI activity among board-certified internal medicine physicians between correct responses, incorrect responses, and guessing (random responses) on these validated MCQs. Correct responses were anticipated to be associated with expertise and incorrect responses to be associated with not having expertise on the item. Guessing was determined to be present when the participant voiced phrases during the think-aloud such as “I have no idea” or “I am guessing”. The data from the think-aloud procedure was used to provide substantiating evidence of expertise versus non-expertise on correct versus incorrect items, respectively. We also

explored potential differences in brain activity with think-aloud data consistent with analytic versus non-analytic reasoning, as well as deep (the major principle(s) that underlie the problem or schema) versus superficial problem representation. For example, for a patient with multiple cardiac risk factors who presented with crushing retrosternal chest pain radiating to the jaw, a deep representation would be ischemia whereas a superficial representation would be chest pain. Deep problem representation is associated with expertise in multiple fields outside of medicine.¹¹ We used each of these three independent analyses (correct, non-analytic, deep) because studies suggest that they are all markers of expertise.^{11,15} Finally, we assessed fMRI differences with answering versus reflecting upon answers. This was done to explore a longstanding debate regarding thinking and metacognition (thinking about one's thinking or awareness of one's reasoning process)¹¹ and the extent to which metacognition might bias think-aloud protocols, (which in turn) explicitly focus on "unfiltered" thoughts related to task performance. In this light, the central assumption of think-aloud protocol analysis is that it is possible to instruct participants to verbalize their thoughts in a manner that does not alter the sequence and content of thoughts mediating the completion of a task.¹¹ The assumption is that metacognition during a think-aloud does not change the sequence and content of thoughts. If this assumption is correct, one would expect that fMRI patterns of activation would be the same with answering an item and thinking-aloud about the item.

We conducted this study with the expectation of finding functional neuroimaging differences related to expertise. In accordance with one tenet of control network theory, we expected that being correct, non-analytic, and using deep problem representation would each lead to more focal fMRI activation, which may be consistent with expert performance (non-analytic reasoning). However, given emerging work,^{10,28} we anticipated that clinical reasoning expertise might also involve the prefrontal cortex and precuneus. Regarding being incorrect, analytic and using superficial problem representation, we expected more general fMRI activation that is thought to be consistent with non-expert performance (analytic reasoning) and the activation of control network theory areas. To examine these questions, we looked for differences across items within individuals (intra-individual differences) and between groups of participants (inter-individual differences). In this way, participants could serve as both an expert as well as a non-expert, provided that the items were challenging enough that the participants did not know the answer.

Methods

Participants

We invited board certified internal medicine physicians (general internists and the subspecialists in internal medicine) to participate in our study. Participants are staff physicians at the Uniformed Services University (USU), Walter Reed Army Medical Center (WRAMC), or National Naval Medical Center (NNMC). Board-certified internists were enrolled in this study after completing informed consent. The following were exclusion criteria: the presence of surgical metal devices or other pieces of metal, an inability to complete an fMRI due to anxiety or claustrophobia, taking calcium channel blockers (which can affect regional

blood flow), being in practice for more than 20 years, or pregnancy. Our study was approved by the Institutional Review Boards at the USU and WRAMC.

Measurements

Multiple-choice questions (MCQs)

We used validated multiple-choice questions (MCQs) from the American Board of Internal Medicine (ABIM) and National Board of Medical Examiners (NBME) as “gold standards” for expertise in diagnostic reasoning. These organizations are responsible for licensing physicians in the United States, and these questions undergo rigorous review. In short, these questions are tested on thousands of subjects and are then modified, as needed, to optimize their psychometrics. We selected questions from two fields in medicine: cardiology and rheumatology. We limited the number of fields in an attempt to dampen the potential effect of context specificity on our results. Also, higher-order MCQs were selected to require the integration and synthesis of data. Participants answered a total of 32 questions based on previous statistical analysis: 8 easy and 8 hard NBME items (United States Medical Licensing Exam Step 2 Clinical Knowledge items); 8 easy and 8 hard ABIM items (Maintenance of Certification MCQ). Questions were selected that fit on a single computer screen and the questions contained words only (i.e. no chest x-rays or other images). The items selected also had favorable discrimination when previously administered, such that high performing individuals answer the hard questions correctly and only the lowest performing individuals miss the easy items. These MCQs represent a “gold standard” for assessing expertise and have robust psychometric data. The opportunity for participants to push handheld buttons for answer options “A” to “E” also made the MCQs ideal for use in an fMRI scanner, which precludes speaking as jaw motion impairs image interpretation.

fMRI measurement

Subjects were scanned on a 3T 750 MRI scanner (General Electric, Milwaukee, WI) with a 32-channel head coil. Acquisitions were performed using an echo-planar imaging (EPI) sequence of 40 contiguous sagittal slices per brain volume (TR = 2000 ms, TE = 25ms, flip angle = 60°, slice thickness = 4.0 mm). In-plane resolution was 3.75 x 3.75 mm (64 x 64 voxels). An fMRI task presentation of the 32 questions was created using E-Prime software (Psychology Software Tools, Inc.) and displayed via a goggle system (Nordic NeuroLab Inc., Milwaukee, WI) while each participant was in the fMRI scanner. The questions were presented in a random order for each subject over the course of four fMRI acquisition runs, with 8 questions per run. The exact length of each run varied, depending on the amount of time each subject took to progress through the reading and answering phases of each question. The mean run length (\pm standard deviation) was 392 ± 62 sec. Subjects received pre-training to acquaint them with the tasks of answering MCQs in the scanner, to be comfortable with the method and layout of question presentation, and in the correct use of the buttons corresponding to answer options A to E. Participants were also trained in the think-aloud procedure,^{11,29} for which a modified version was used in the reflection phase for items

in the scanner; that is, participants were asked to think-aloud without speaking, as jaw motion would affect images. During the same imaging session, a high-resolution T1-weighted image was acquired for anatomical reference (three dimensional GRE; TR = 6.6 ms, TE = 2.5 ms, flip angle = 12 degrees). This image consisted of 312 sagittal slices with a slice thickness of 0.6 mm and an in-plane resolution of 0.468 x 0.468 mm (512x512 voxels). A B0 field map was also acquired to correct for field inhomogeneity in the fMRI EPI images.

For each MCQ, participants were given up to 60 seconds to read the question, and each question was projected in three phases. In the first phase, the question stem appeared (reading phase), ending with “what is the most likely diagnosis?” or a related diagnostic question, but not displaying answer options “A” to “E”. Each participant was given 60 seconds during the reading phase to push any button to move on to the answer options (the second or answering phase). Subjects were then given 7 seconds to choose an answer option using the finger response buttons. The final phase (reflection phase) was then entered, where participants were instructed to think, without speaking, about how they arrived at the diagnosis (“how did you establish the diagnosis for this item?”). This phase lasted 14 seconds. There was not a formal baseline period between questions.

Think-aloud protocol

Participants completed a formal think-aloud procedure for all items immediately following completion of the items in the scanner. Participants received formal training on the think-aloud procedure before entering the scanner, consistent with published methodology.^{11,29} More specifically, participants were asked to describe in words their process of answering the questions (to “think-aloud”) while reviewing each item just answered in the fMRI scanner. Participants were handed a printout of items with “A” to “E” answer options; however, individual participant’s answers were not included. Each think-aloud procedure was performed by a trained investigator, and audio-taped.

Data analysis

All fMRI data were processed using the AFNI software package. Image preprocessing included: removal of the first three volumes (6 sec) from each 4D time series, slice-time correction, motion correction, correction for inhomogeneities in the B0 field, registration to the T1 anatomical image, smoothing with an 8 mm full-width-half-max Gaussian kernel, and conversion of voxel values to percent-change-from-mean rather than absolute intensity. Of the 18 subjects, 1 subject’s fMRI data set was excluded from further analysis due to motion artifact while in the scanner.

Activation analysis was performed using a general linear model (GLM) approach concatenating the four data sets for each subject. Hemodynamic response estimates were modeled for each of the three question phases (read, answer, reflect). The “read” and “answer” times varied from question to question and were modeled with a gamma-variate function with variable duration and variable amplitude. The “reflection” time was constant at 14 seconds and was modeled with a non-variable gamma-variate. The GLM analysis was used to determine the significance of these model time-courses within each voxel. Time-courses

associated with preprocessing motion correction parameters were also included in the GLM regression analysis to regress out any further image intensity changes due to motion. As we sought to capture the construct of reasoning, we compared phases (reading, writing, reflecting) to elicit task specific findings. Thus, the “reading” time periods were treated as a baseline in most of our analyses--most fMRI analysis results are the result of differences between relative activation between “reading” and “answer” or “reflection” phases. We also specifically compared “answer” and “reflection” phases.

As a first level of analysis, contrasts were made between the “answer” and “reflection” significance estimates in each voxel for each subject (to determine which regions, if any, showed activation differences between these two phases). A group analysis comparing this contrast across all 17 subjects was then performed using a linear mixed-effects modeling approach. Each subject’s response to each of the 32 questions was categorized based on four characteristics: correct vs. incorrect, guess vs. non-guess, analytic vs. non-analytic, deep vs. superficial; the latter three categories were arrived at by analyzing audio-taped transcripts of the post-scan think-aloud procedure. Eight group-wide contrasts were run on a voxel-wise basis to determine if these distinctions correlated with differences in the “answer” or “reflection” response (contrasts were run between “answer, correct” and “answer, incorrect”, “reflection, correct” and “reflection, incorrect”, “answer, deep” and “answer, superficial”, etc.). These contrasts were set up to determine if there were any brain regions that showed different activation responses based on response categorizations. For all group analyses, age and gender were included as covariates.

For determining analytic versus non-analytic, deep versus superficial problem representation, and guessing versus not guessing, three coders independently reviewed the audio-taped think-aloud protocols of participants and then discussed results. Agreement on categorization was achieved by consensus. Two coders then reviewed the transcripts in their entirety. Each of these three scores was coded and analyzed independently.

Results

A total of 17 board certified internists were included in our analysis (of which one participant’s data set was dropped due to excessive motion in the scanner). The mean number of correct responses was 18.5/32 questions (range 15-25). There were 15 men and 2 women. Mean age was 39.5±7 (range=32-51 years). Participant specialties were: 9 general internal medicine, 3 cardiology, 2 pulmonary/critical care, 1 gastroenterology, 1 infectious diseases, and 1 hematology-oncology. All participants were board certified in internal medicine, and all but one in their respective subspecialty (a cardiology fellow). Significant fMRI activation locations are shown in Table 1.

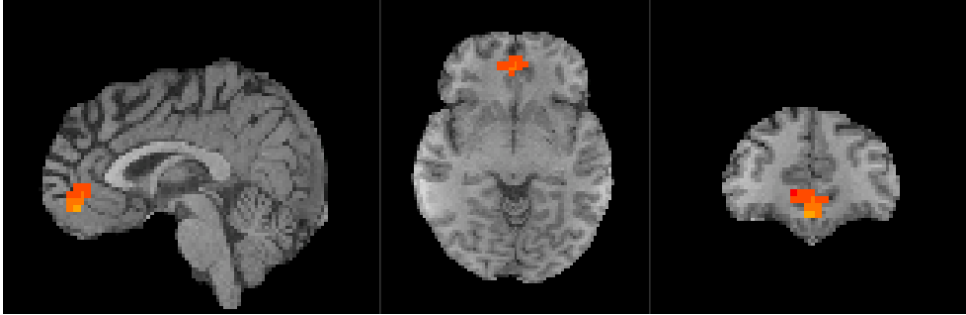
Table 1: Significant Contrasts (Comparisons) with Functional Neuroimaging*

Contrast	Location	MNI Coordinates			Corrected p value
		x	y	z	
incorrect vs. correct	medial PFC	-2	-47	-4	< 0.01
guess vs. non-guess	medial frontal gyrus	-1	1	47	< 0.01
	precuneus	1	50	49	< 0.01
	L mid temporal gyrus	57	55	7	< 0.01
answer vs. reflection	occipital cortex	-6	71	6	< 0.01
	L basal ganglia	18	-14	6	< 0.01
	R basal ganglia	-14	-6	2	< 0.01
	R PFC	40	-26	22	< 0.01
	L PFC	-52	-8	28	< 0.01
	L cerebellum	19	52	21	< 0.01
	R cerebellum	-26	62	-35	< 0.01
	precuneus	2	48	32	< 0.01

*= Contrast or comparison; incorrect vs. correct and guess vs. non-guess (obtained from think-aloud analysis). PFC=prefrontal cortex; MNI=molecular neuroimaging; L=left and R=right

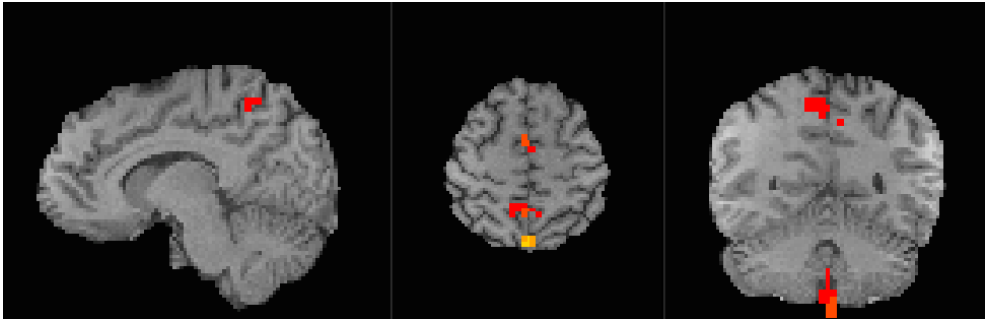
Examining neuroimaging activation locations for correct versus incorrect answers revealed one statistically significant difference--greater activation of the prefrontal cortex with answering incorrectly versus correctly ($p<0.01$; Figure 1). There were no other significant differences in brain activation when making comparisons within individuals or across subjects. There were no activation differences in the reflection phase with correct versus incorrect responses.

Figure 1: A region in the medial PFC showed greater activation for incorrect versus correct answers during the “answer” phase of each multiple-choice question.



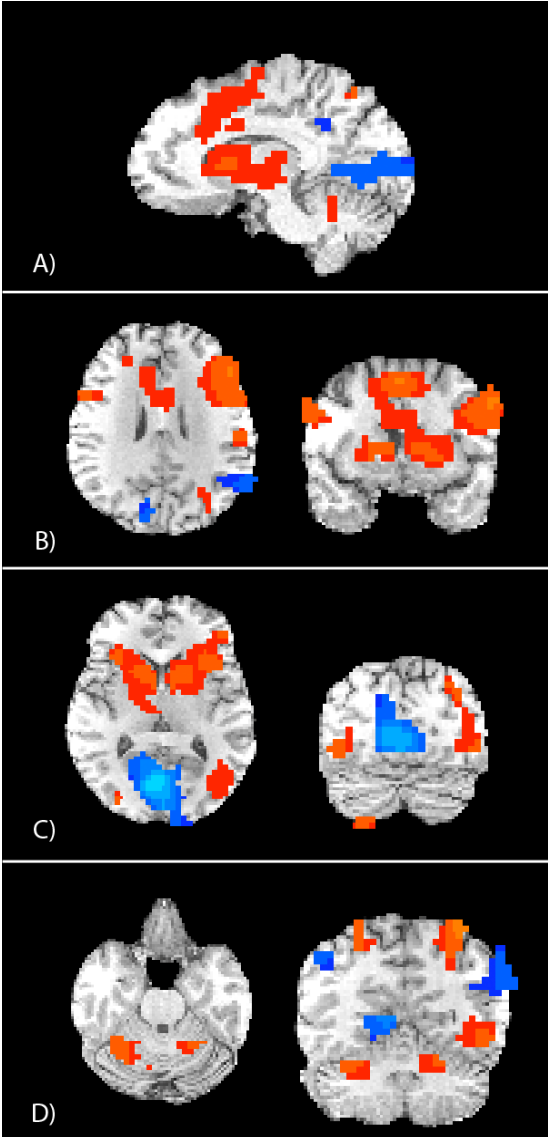
Guessing was determined using the formal think-aloud data. Two brain areas were significantly less activated ($p < 0.01$) with guessing versus not guessing during the answering phase: bilateral precuneus (PCC) and left mid temporal gyrus (Figure 2). There were no significant differences in guessing versus not guessing during the analysis of the reflection phase.

Figure 2: Providing answers that were the result of guessing produced less activation in the precuneus during the “answer” phase of each multiple-choice question than did providing answers that did not require guessing.



We also compared brain activity associated with answering items versus reflecting on items. There was more activation with answering for the following areas: motor cortex (which was expected as participants pushed the buttons in the answering phase but not in the reflection phase), bilateral prefrontal cortex, bilateral cerebellum, and the caudate/putamen (Figure 3). Brain activity was greater with reflection versus answering for the following areas (all $p < 0.01$; Figure 3): occipital cortex, precuneus (PCC).

Figure 3: Several regions, including the motor cortex, bilateral prefrontal cortex, bilateral cerebellum, and the basal ganglia showed more activation while answering the multiple-choice questions than while reflecting upon arriving at a chosen answer. Regions of the occipital cortex and precuneus showed greater activation while reflecting than while answering.



There were no significant differences in the fMRI activation with answering or reflecting phases for comparing think-aloud coded analytic versus non-analytic responses or deep versus superficial representation. Notably, the hippocampus, which is frequently associated with memory tasks, was not significantly activated for any of the comparisons we performed.

Discussion

Understanding the complex cognitive processes underlying clinical reasoning is essential for promoting high-quality patient care. Prior work in measuring clinical reasoning has been challenging due to the impact of context specificity on the clinical reasoning process and the lack of precise instruments to capture this context specificity. To address these challenges, we explored diagnostic reasoning expertise using functional neuroimaging, think-aloud protocols, and validated high-stakes MCQs. We then examined the functional neuroimaging evidence to enhance our understanding of current educational theory regarding diagnostic reasoning, which we believe in turn provides a number of future practical implications for future research.

Our study yields potentially important findings that build on the fMRI and educational theory literature. Furthermore, our findings show evidence of validity because they replicate some results from fMRI research in fields outside of medicine.²⁶

Prefrontal cortex activation

We identified statistically significant differences involving the prefrontal cortex and the basal ganglia for answering correctly versus incorrectly and for answering questions versus reflecting. The prefrontal cortex is widely believed to be involved in executive functions, such as decision making,^{30,31} which is consistent with our finding of prefrontal activation with correct versus incorrect answers, as well as differences with answering versus reflecting on answers in general. The prefrontal cortex has also been implicated as a potential area for working memory processing.^{20,25} Notably, incorrect answers resulted in more prefrontal cortex activation than correct answers. Simply put, the cognitive regions of the brain may have had to work harder to address more difficult questions. This area likely serves as a component of the control network theory for analytic clinical reasoning. Additionally, the prefrontal cortex may represent an area that is used for working memory and clinical reasoning. Activation of a greater amount of working memory may represent an important fMRI distinction between analytic and non-analytic reasoning. The prefrontal cortex has also been recently cited as a brain region common to different domains of wisdom, which includes a component of rational decision-making and insight,¹⁰ adding further validity to our findings of the use of the prefrontal cortex in internal medicine expertise.

Caudate and precuneus

The caudate and putamen are structures in the basal ganglia that have recently been implicated in playing roles in learning and memory. We found increased activation in these areas

with answering versus reflecting on items, which is consistent with prior research.³² Similarly, recent fMRI studies²⁸ implicate the precuneus in self-referential goal-directed actions, as well as memory retrieval. These functions are consistent with our finding of increased activation when not guessing on an item and with reflection on answered items. These patterns of activation would not be expected by control network theory and the precuneus may represent a focus of expertise (non-analytic reasoning or both analytic and non-analytic reasoning as studies suggest experts can use both systems). Additionally, the striatal regions are also rich in dopaminergic neurons, which have been shown to be central to reward mechanisms.³³ The increased activation in these regions suggests that the act of answering the question produces a reward signal, which is likely in this group of highly motivated individuals accustomed to being tested. These differences with answering versus reflecting have potential implications for think-aloud analysis which could be further explored.

Occipital cortex

Another significant finding is that the occipital cortex was more activated during reflection versus answering items. One potential explanation is that our experts were visualizing actual clinical scenarios with reflecting on their answers. Another potential explanation is that different background colors for items were used as a cue to participants that they were in the reflection period of each MCQ. Different colors have been shown to trigger various changes in brain activation.³⁴ Also, the finding of greater mid occipital gyrus activation while answering versus reflecting may be due to its potential role in working memory.^{20,25}

Temporal lobe

We observed accentuated activation of the medial temporal lobe when not guessing; like the precuneus, this area may represent a focus for clinical reasoning expertise. It is interesting to note that several studies have shown that damage to the medial temporal lobe is associated with an impaired ability to learn and remember new information (i.e. anterograde amnesia). This finding has been replicated in other studies and the parietal-temporal area has been implicated in reasoning about meaningful content.²⁷ This finding suggests that participants were activating learning areas when not guessing and perhaps were giving up on the item with guessing, resulting in less medial temporal lobe activation.

Cerebellum

Our finding of cerebellum activation with answering items is also consistent with emerging work implicating the cerebellum as a working memory area.^{35,36} Similar to other fields,^{20,27} and as noted above, we found several areas that may be involved in working memory with clinical reasoning tasks. The potential multi-neuroanatomic foci of working memory may be a characteristic of clinical reasoning expertise. Expertise may represent a broad blend of cognitive functions and thus lead to a wide array of activation of working memory areas.

The finding of few differences between answering and reflecting on answers, by functional neuroimaging supports the use of think-aloud protocols for measuring cognitive

processes. The previously suggested bias associated with “thinking about one’s thinking” for purposes of these protocols may be inaccurate.^{11,29} Our findings have important implications for adding tools that explore cognition, such as think-alouds, to tools that measure behaviors, such as post encounter forms in Objective Structured Clinical Exams or answers to MCQs. Indeed, our fMRI finding of few differences with answering versus reflecting would support increasing the use of think-aloud protocols alone or in combination with other measurements when exploring clinical reasoning. This is encouraging as think-aloud protocols are a feasible and relatively low cost means to assess clinical reasoning.

Several of our analyses revealed no significant differences. We did not identify functional neuroimaging differences for think-aloud categorization of analytic versus non-analytic reasoning or deep versus superficial problem representation, and we found few differences with correct versus incorrect and guessing versus non-guessing. Expertise in internal medicine does not appear to have a uniform pattern of fMRI activation. There may be several reasons for this. For example, the brain processes corresponding to analytic and non-analytic theories may be more complex and subject to wide inter-individual differences, or the changes in regional perfusion may have been too small to detect with our methods and our small exploratory sample. Experts may use both analytic and non-analytic processes with solving the same vignette. It is also possible that the differences may be nonlinear (or the outcome is more than the sum of the parts, interacting in complex and dynamic ways) and thus our general linear model, which does not account for such interactions, would not detect differences. Also, our inability to detect significant differences may have been due to conducting the think-aloud analysis after a short period of delay (completing questions in the scanner) which could introduce recall bias. However, given that participants rarely (<5%) endorsed not remembering what they answered in the scanner, this adds validity evidence to our procedure. Work in the fMRI field has suggested nonlinearity with other brain processes.³⁷⁻⁴⁰ The notion of nonlinearity with complex tasks is also consistent with emerging educational psychology theories such as situated cognition and ecological psychology.⁴¹⁻⁴³ Another possibility refers to the problem of poor self-assessment.^{44,45} If our brains are using the same (or nearly the same) activation patterns for the above processes, this tight coupling could provide insight into why physicians may fail to recognize when they are wrong—the patterns of activation may be too tightly coupled to recognize distinctions.

Our study has several limitations. First, our study may have had insufficient power to detect fMRI differences. Second, we may have failed to find fMRI differences due to range restriction from sampling only experts, though our performance means and standard deviations for items suggest that this did not play a large role. However, future work should consider comparing experts to novices. Third, new paradigms for fMRI analysis could be considered that might enlighten future work, such as attempting to formally conduct the think-aloud procedure in the fMRI scanner, and adjusting the time for each phase (reading, answering, reflecting). Fourth, participant performance may have been affected by answering items in the scanner. Though we provided training prior to entering the scanner and no participant reported significant claustrophobia, answering items with restricted movement in the scanner may not be the same as answering items in a testing center. This finding would be consistent with the notion of embodied cognition; that is, the idea that cognition is often paired with movement and so the inability to move in a fMRI scanner may have

negatively impacted thinking and performance.^{41,46} Fifth, the reasoning associated with answering MCQs may be quite different than reasoning applied to seeing actual patients. Future work could consider having participants view videotapes of patient-physician encounters. Sixth, we did not include images in the items such as chest x-rays or ECGs; these images could impact findings. Seventh, fMRI studies require comparisons, and perhaps our comparisons between 3 active states (reading versus reflecting, reading versus answering, and answering versus reflecting), without an explicit background “resting” state limited our findings. We do believe, however, that a strength of our study was having participants think-aloud immediately following answering the item (reflection time) in order to accurately capture the functional neuroimaging of the construct of metacognition. Moreover, the time taken for think-aloud following the fMRI procedure with reviewing all the items answered in the scanner, was consistent with (not significantly different from, data not shown) time to answer questions in the scanner, providing some validity evidence to our modified think-aloud procedure in the scanner. An eighth limitation is the lack of a formal rest or inactive period to compare with the task phases; as we sought to capture the construct of reasoning, we believe that comparing phases (reading, writing, reflecting) would result in more meaningful, task specific findings, understanding that there can be features of answering or reflecting while being engaged in the process of reading an item. Ninth, we did not explore fMRI patterns of board-certified internists (experts) who have been in practice for more than 20 years and our findings also may not generalize to other specialties and/or those who have been in practice longer. Tenth, we did not control for other situational factors, such as fatigue or well-being, which could have impacted our results. Work outside of medicine would suggest that these factors are important.^{41-43,47}

This is the first study to utilize fMRI to investigate clinical reasoning expertise. Consequently, this study has several implications for future research in medical education. First, while brain processes appear to be more complex than suggested by control network theory, we did find evidence to support analytic and non-analytic reasoning theory, and we built upon recent work that has suggested additional regions of the brain that may be important to clinical reasoning such as the caudate, occipital lobe and the putamen. The lack of hippocampal activation is consistent with emerging work that memory traces of repeated associations, which, as would be expected in experts, are consolidated in neural networks outside of the hippocampus.⁴⁸ Alternatively, lack of hippocampus findings could be related to lack of an explicit background resting period, as it is possible that subjects were generally accessing hippocampal memory for all three of the measured phases (reading, response, reflection). We also found that experts may use multiple areas for working memory that has been shown in studies of simple tasks outside of medicine.^{20,25} There does not appear to be one consistent fMRI activation pattern of internal medicine expertise. Obtaining measurable functional neuroimaging differences may require larger sample sizes or contrasting groups at different levels of training, such as novices versus experts.

There are several potential implications of our findings for medical education theory and research. First, we found functional neuroimaging evidence to support dual processing, but the distinction between analytic and non-analytic processes proved to be more complicated than originally thought. Alternatively, expert physicians may use multiple methods while solving a given problem. Work has previously suggested that analytic processing, by its nature of being effortful, would be slow; this may not always be the case—there may be

times that analytic processing is a bit faster and so well-practiced that it requires minimal effort, effort that the participant may not even consciously perceive. In other words, analytic and non-analytic theories may over-simplify what actually happens in the expert. Second, working memory for clinical reasoning may be diffuse and given the areas of activation, likely multi-modal in experts. This work adds some evidence to cognitive load theory's notion of multi-modal processing and the potential educational benefits of teaching with multiple modes to decrease cognitive load.⁴⁹⁻⁵¹ Third, clinical reasoning in experts appears to be complex (and potentially nonlinear) and, as such, is likely subject to large inter-individual differences; consequently, educational assessment should not be assumed to be a "one size fits all" process. Our findings would support additional work in clinical reasoning functional neuroimaging--identifying patterns of activation could potentially assist with teaching and remediating this essential competence for medicine and could help better integrate existing theories on clinical reasoning and expertise. For example, comparing experts' findings with individuals who have failed to meet board certification or recertification standards to see if functional neuroimaging differences exist. Current behavioral measures such as answering MCQs or completing post-encounter forms do not appear to be optimal measures of this complex mental process. Fourth, given the small differences between answering and reflection on answers, think-aloud protocols may be more valuable than previously thought, especially when subjects receive formal training on the procedure. This supports the use of tools that go beyond observable behaviors by capturing internal mental processes.

Our findings support emerging fMRI literature,^{10,52} indicating that fMRI may help educators with assessing complex cognitive processes, such as diagnostic reasoning. Ultimately, we believe the measurement approach of combining neuroimaging and think-aloud data with more traditional assessment measures offers great potential for advancing our understanding of diagnostic reasoning.

References

1. Cervero RM. *Effective Continuing Education For Professionals*. San Francisco (CA): Jossey-Bass; 1988.
2. Higgs J and Jones MA. Clinical decision making and multiple problem spaces. In: *Clinical reasoning in the Health Professions*, 3rd ed. Higgs J, Jones MA, Loftus S, and Christensen N, eds. Oxford: Butterworth Heinemann; 2008.
3. Ark TK, Lee R Brooks LR, and Eva KW. The benefits of flexibility: the pedagogical value of instructions to adopt multifaceted diagnostic reasoning strategies. *Med Educ*. 2007;41:281-287.
4. Norman G. Research in clinical reasoning: past history and current trends. *Med Educ*. 2005;39:418-427.
5. Elstein AS, Shulman LS, Sprafka SA. Medical problem solving: a ten-year retrospective. *Eval Health Prof*. 1990;13:5-36.
6. van der Vleuten CP, van Luyk SJ, Swanson DB. Reliability (generalizability) of the Maastricht Skills Test. *Res Med Educ*. 1988;27:228-33.
7. Schuwirth LW, van der Vleuten CP. The use of clinical simulations in assessment. *Med Educ*. 2003;37:65-71.
8. Eva K. Exploring the etiology of content specificity: Factors influencing analogic transfer and problem solving. *Acad Med*. 1998;73:S1-S5.
9. Chaiklin S and Lave J. *Understanding Practice: Perspectives on Activity and Context*. New York: Cambridge University Press; 1993.
10. Jeste DV, Harris JC. Wisdom—A neuroscience perspective. *JAMA*. 2010;304:1602-1603.
11. Ericsson KA, Charness N, Feltovich P, Hoffman RR, eds. *The Cambridge Handbook of Expertise and Expert Performance*. New York (NY): Cambridge University Press; 2006.
12. Gonsalves BD and Cohen NJ. Brain imaging, cognitive processes, and brain networks. *Perspect Psychol Sci*. 2010;5:744-750.
13. Ruiter DJ, van Kesteren MTR, Fernandez G. How to achieve synergy between medical education and cognitive neuroscience? An exercise on prior knowledge in understanding. *Adv in Health Sci Educ*. 2010.
14. Goswami, U. Neuroscience and education: from research to practice? *Nat Rev. Neurosci*. 2006;7:406-411.
15. Eva K. What every teacher needs to know about clinical reasoning. *Med Educ*. 2004;39:98-106.
16. Ericsson KA, Kintsch W. Long term working memory. *Psychol Rev*. 1995;102:211-45.
17. Norman GR, Tugwell P, Feightner JW, Muzzin LJ, Jacoby LL. Knowledge and clinical problem solving. *Med Educ*. 1995;19:344-356.
18. Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise: theory implications. *Acad Med*. 1990;65:611-621.
19. Shell DF, Brooks DW, Trainin G, Wilson KM, Kauffman DF, Herr LM, eds. *The Unified Learning Model: How Motivational, Cognitive, and Neurobiological Sciences Inform Best Teaching Practices*. London: Springer; 2010.
20. Baddeley AD. *Working Memory*. Oxford: Clarendon Press; 1986.
21. Hill NM and Schneider W. Changes in neural activation related to dual-task practice: evidence for a domain general learning network. 2005. Poster presented at the 11th annual meeting of Human Brain Mapping, Toronto, ON.
22. Kelly AMC, Garavan H. Human functional neuroimaging of brain changes associated with practice. *Cerebral Cortex*. 2005;15:1089-1102.
23. Poldrack, RA. Imaging brain plasticity: conceptual and methodological issues. *Neuroimage*. 2000;12:1-13.
24. Schneider W, Chein JN. Controlled and automatic processing: behavior, theory, and biological mechanisms. *Cognitive Sci*. 2003;27:525-559.
25. Baddeley A. Working memory. *Science*;1992;255:556-559.
26. Chein JM, Schneider W. Neuroimaging studies of practice-related change: fMRI and meta-analytic evidence of a domain-general control network for learning. *Brain Res Cogn Brain Res*. 2005;25:607-623.
27. Anderson JL. *Cognitive Psychology and its Implications*, 7th ed. New York (NY): Worth Publishers;2010.
28. Cavanna AE, Trimble MR. The precuneus: a review of its functional anatomy and behavioral correlates. *Brain*. 2006;129:564-583.
29. Ericsson KA, Simon HA. *Protocol Analysis: Verbal Reports as Data*. Cambridge: MIT Press; 1984.
30. Stuss D and Knight RT, eds. *The Frontal Lobes*. New York: Oxford University Press; 2002.
31. Shallice T. *From Neuropsychology to Mental Structure*. 1988. Cambridge: Cambridge University Press;1988.
32. Graybiel AM. The basal ganglia: learning new tricks and loving it. *Curr Opin Neurobiol* 2005;15:638-644.
33. Koeppe MJ, Gunn RN, Lawrence AD, Cunningham VJ, Dagher A, Jones T, Brooks DJ, Bench CJ, and Grasby PM. Evidence for striatal dopamine release during a video game. *Nature*.1988;393:266-8.

34. Beauchamp MS, Haxby JV, Jennings JE, DeYoe EA. An fMRI version of the Farnsworth-munsell 100-hue test reveals multiple color-selective areas in human ventral occipitotemporal cortex. *Cerebral Cortex*. 1999;257-263.
35. Desmond JE, Chen SH, DeRosa E, Pryor MR, Pfefferbaum A, Sullivan EV. Increased frontocerebellar activation in alcoholics during verbal working memory: an fMRI study. *Neuroimage*. 2003;19:1510-20.
36. Courtney SM, Ungerleider LG, Keil K, and Haxby JV. Object and spatial visual working memory activate separate neural systems in human cortex. *Cerebral Cortex*. 1996;6:39-49.
37. Friston KJ, Mechelli A, Turner R and Price CJ. Nonlinear responses in fMRI: the balloon model, volterra kernels, and other hemodynamics. *NeuroImage*. 2002;12:466-477.
38. Lauritzen M and Gold L. Brain function and neurophysiological correlates of signals used in functional neuroimaging. *J Neurosci*. 2003;23:3972-3980.
39. Janz C, Heinrich SP, Kornmayer J, Bach M, Hennig J. Coupling of neural activity and BOLD fMRI response: new insights by combination of fMRI and VEP experiments in transition from single events to continuous stimulation. *Magnet Reson Med*. 2001;46:482-486.
40. Kelso JAS. *Dynamic Patterns: The Self-organization of Brain and Behavior*. Cambridge (MA): MIT Press; 1995.
41. Robbins P and Aydede M. *The Cambridge Handbook of Situated Cognition*. New York (NY): Cambridge University Press; 2009.
42. Bredo E. Reconstructing educational psychology: Situated Cognition and Deweyian Pragmatism. *Educ Psychol*. 1994;29:23-35.
43. Brown J. S., Collins, A. and Duguid, P. Situated cognition and the culture of learning. *Educational Researcher*. 1989;18:32-42.
44. Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. *Acad Med*. 2005;80:S46-S54.
45. Davis DA, Mazmanian PE, Fordis M, Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: A systematic review. *JAMA*. 2006;296:1094-1102.
46. Anderson ML. Embodied cognition: a field guide. *Artif Intell*. 2003;149:91-130.
47. Shaw RE and Bransford J, eds. *Perceiving, Acting and Knowing: Toward an Ecological Psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1977.
48. Miyashita Y. Cognitive memory: cellular and network machineries and their top down control. *Science*. 2004;306:435-440.
49. Van Merriënboer JJG, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ*. 2010;44:85-93.
50. Sweller J, Chandler P. Why some material is difficult to learn. *Cognition Instruct*. 1994;12:185-233.
51. van Merriënboer JJG, Sweller J. Cognitive load theory and complex learning: recent developments and future directions. *Educ Psychol Rev*. 2005;17:147-177.
52. Riess H. Empathy in Medicine—A neurobiological perspective. *JAMA*. 2010;304:1604-1605.

Chapter 11

Discussion

General Discussion

The purpose of this thesis was to explore the influence of contextual factors on clinical reasoning during the clinical encounter. More specifically, the overarching research question was: do selected contextual factors impact clinical reasoning performance, both in expert practice and during instruction of trainees? A main focus of this thesis was improving our understanding of context specificity as it pertains to clinical reasoning performance. This understanding includes exploring (unraveling) the problem of context specificity germane to expert performance, instructional methods, and methods of assessment.

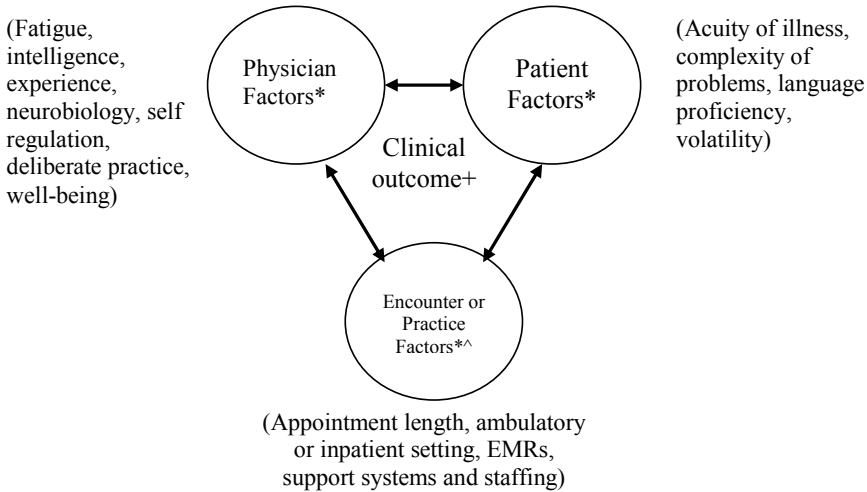
The emphasis of this thesis was on diagnostic reasoning, though findings emerged that are also relevant to therapeutic clinical reasoning. In this thesis, we did not attempt to measure reasoning from a standard psychometric sense; instead, we attempted to capture clinical reasoning in more naturalistic (*in vivo*) settings. In this light, the importance of theoretical frameworks, and the extent to which they might provide a lens for future investigation, was emphasized. Indeed, an overarching implicit goal of this thesis was to construct a “roadmap” or “compass” for the future investigation of the influence of contextual factors on clinical reasoning. In other words, we hoped the investigations would yield evidence to contribute to these theoretical frameworks, by providing evidence for the theories and/or possible revisions, as well as by helping to inform additional important avenues for future investigation in regards to understanding the influence of context on clinical reasoning. In this section of the thesis, we return to our overarching theoretical frameworks and then discuss study findings for each research question, general conclusions, limitations, recommendations for future research, and, finally, implications.

Overarching theoretical frameworks

Two unifying theoretical frameworks were used in our investigations of the impact of contextual factors on clinical reasoning performance--situated cognition and ecological psychology. These theories argue that it is not possible to separate the participants' reasoning in a clinical encounter from the environment in which it occurs.

Situated cognition provides a unique perspective for studying the influence of context on clinical reasoning by considering the individual factors in the encounter (patient, physician, practice) and their interactions.¹⁻⁴ Situated cognition's perspective facilitates both a way to view the dynamic interplay of component parts that *emerge* or *evolve* in the overall clinical encounter as well as a means to diagrammatically understand the situation dependent (or situated) nature of clinical reasoning (Figure 1).

Figure 1: Situated Cognition, Ecological Psychology, Cognitive Load Theory and Understanding the Impact of Contextual Factors in the Clinical Encounter



*=factors that interact as shown by arrows; parentheses next to each factor are examples of elements of each factor
Situating cognition takes the approach of the individual and the environment and all of the above interactions can and do influence the outcome (patient care) in such a model

+clinical outcome is dependent upon these 3 factors, their interactions, and possibly other inputs.

Bidirectional arrows (interactions) are governed by: participants' multiple goals, affordances/effectivities, cognitive load, authenticity

^=encounter or practice factors refer to elements pertaining to the environment or setting for the encounter

Indeed, we believed that this theory would be informative to our investigations, as the situation-dependent nature of clinical reasoning is central to the problem of context specificity. Situated cognition proposes a way to view clinical reasoning in both the larger social (participants) and physical (environment or setting) context. Therefore, the chapters in this thesis consistently applied this three-factor approach (patient, physician, and practice or encounter) with investigating the influence of contextual factors on clinical reasoning.

Ecological psychology supports the tenets of situated cognition but it does so from a slightly different perspective, focusing on the individual agent (participant) and environment interactions.⁵⁻⁷ In other words, ecological psychology provides a perspective for what *drives* the bidirectional arrows in Figure 1 between participant groups as well as participant groups and the environment. As such, we believe that ecological psychology adds to situated cognition in at least two important ways, and, therefore, this theory was also consistently applied within the chapters in this thesis. First, ecological psychology, proposes that clinical reasoning *emerges* through these interactions (bidirectional arrows in Figure 1) based on affordances (what the environment provides to stakeholders or opportunities for action) and effectivities (what the specific stakeholder brings to environment based on their abilities). This proposition was specifically addressed in two chapters (chapter 4 and 7). Second, ecological psychology proposes that participants often have several goals in an encounter and these goals can, and do, change in their priority; indeed goals are believed to

be a primary driving behavior in these interactions. This idea was supported by our qualitative study (chapter 7), which sought improve our understanding of how contextual factors may influence clinical reasoning in experts. These important dyads: affordances/effectivities and evolving goals of participants, give a glimpse into why we believe context is nested, or highly dependent upon participant and practice inputs (Figure 1), which may lead to nonlinearity and even complexity, manifest as deterministic chaos in certain situations.⁸⁻¹⁰ These ideas challenge the current classic linear methodological approaches to studying the problem of context specificity and clinical reasoning.

A third theory that emerged during this program of research, and which played a major role in several of the chapters of this thesis, is cognitive load theory (CLT). CLT recognizes the limits of our working memory capacity and the importance of long-term memory schemas and automation. CLT therefore provides a unique framework for investigations into cognitive processes,¹¹⁻¹³ such as clinical reasoning. CLT informed several of our investigations, adding an important additional theoretical view to the two overarching theories described above. More specifically, situated cognition and ecological psychology stipulate that reasoning emerges through interactions of participants and their environment (Figure 1 and ecological psychology's concept of dyads alluded to above). An important principle of CLT is the concept of *element interactivity*. Although the information that physicians must process varies on many dimensions, the extent to which relevant informational elements interact has been found to be a critical feature in prior CLT studies.¹¹⁻¹³ Information can be thought of as varying on a continuum from low to high in element interactivity. Each piece of low-element interactivity information can be processed and understood individually, without much consideration of any other information elements, and thus is thought to induce little cognitive load. On the other hand, each piece of high-element interactivity information requires consideration of other elements to be processed and understood and thus can greatly increase demands on cognitive load.¹¹⁻¹³ The principle of element interactivity from CLT provides important insight into why we believe context is nested and the problem of context specificity (Figure 1). Our view of context with multiple factors and their attendant interactions that evolve or emerge, would, in essence, argue for the potential for "higher element interactivity" as well as the likelihood of significant extraneous cognitive load (load that interferes with the task at hand; i.e. through the goal-specific principle).¹¹⁻¹³ In other words, these arguments suggest that CLT may play an important role in understanding context specificity. Extraneous load is likely with all the factors and their potential interactions proposed by the situated cognition model (Figure 1). Importantly, CLT has been used for instructional purposes and not explicitly for exploring cognitive processes in experts and therefore we believe that our thesis findings are germane to advancing the CLT literature.

CLT, combined with the three-factor approach of situated cognition (patient, physician, encounter) and the dyads (interactions) of ecological psychology, provided a useful theoretical framework for the cognitive task of clinical reasoning. In other words, situated cognition and ecological psychology provided the overarching view for exploring context and generating more general prediction, interpretation, and study design principles. CLT helped to inform these overarching theories in terms of two key issues. First, CLT provided an additional hypothesis for how these complex interactions (Figure 1) occur for the cognitive task of clinical reasoning and thus also provided theoretical predictions and potential

implications for investigation findings (chapters 6-9). Second, CLT provided additional insight into what the balancing or competing “forces” entail in these interactions. In sum, CLT provided a “micro-theoretical view” to the “macro-theoretical view” of situated cognition and ecological psychology. Indeed, a recognized shortcoming to situated cognition and ecological psychology is how these inclusive, “bigger picture” theories equate a to specific investigation’s practical implications and next steps.¹⁴ Further, the potential for nonlinearity and even deterministic chaos (results are not completely predictable at the outset; however, results are bounded or fall within a range of possibilities)⁸⁻¹⁰ in complex interactions, such as certain clinical encounters, is plausible. This potential would also be consistent with delicate balances governing the bidirectional interactions (Figure 1 arrows), as the result of ecological psychology’s predicted dyads for the encounter and CLT.

In the next section, we explore the main findings from each of our research questions in greater detail. Our specific research questions in this thesis were:

1. What is the influence of contextual factors in expert practice?
2. Does the authenticity of instructional format impact expertise development in clinical reasoning?
3. Can we improve upon assessment of clinical reasoning by triangulating behavioral measurements with concurrent verbal reports (think-aloud protocols) or with novel instruments such as fMRI scanning?

Research Question 1: What is the influence of contextual factors in expert practice?

In Chapters 2 and 3, we conducted thematic reviews to help conceptualize the potential influence of context on expert practice. These reviews which consulted literature from fields outside of medicine resulted in three important implications:

1. Physician training and practice occur in complex environments (contexts). Context is more than just the setting or additional content. Instead, context in the clinical encounter is a complex system that evolves over time. The outcome (patient care) is driven by interactions and feedback among contextual factors (patient, physician, practice). Chapter 3 reviews how context may impact aging physicians’ clinical reasoning performance in the clinical encounter.
2. The interactions and feedback among participants and setting in the situated cognition framework (Figure 1) would not be expected to be completely predictable. Therefore, the outcomes of clinical encounters could, at times, be nonlinear and potentially even chaotic (deterministic chaos, where results are bounded, meaning that although an outcome cannot be precisely predicted at the outset, there are limits to the possible interactions).⁸⁻¹⁰ This suggests the need to consider other mathematical models when studying the influence of context on clinical reasoning, such as deterministic chaos and hierarchical linear modeling. This is not to say that *all* interactions in the clinical encounter are nonlinear or that many are chaotic; we would argue that *some* may be, and we acknowledge the difficulty in predicting, prospectively, which specific interactions might result in nonlinearity or even deterministic chaos. This issue is not unlike trying to predict when a major storm will arise or when turbulence will be manifest while flying an airplane; even with access to sophisticated mathematical modeling, such predictions are fraught with difficulties and require robust real-time sampling and triangula-

tion of data to improve prediction. Unfortunately, as we have learned from weather catastrophes, chaos may only become apparent in hindsight. Implications from these chapters (2 and 3) include devising studies that explore the potential for nonlinearity or even deterministic chaos in the clinical encounter. Unfortunately, such studies require large numbers of subjects with multiple samples in sequential time using advanced statistical programs.¹⁰ A second implication is the need for accurate and frequent “sampling” in medical education, being mindful that deterministic chaos could theoretically occur with multiple participant groups, goals, affordance and effectivities, and cognitive load issues among participants and the environment.

3. Studying specific, “extreme” forms of individual patient, physician, and practice factors and their interactions could help unravel the influence of context on diagnostic and therapeutic clinical reasoning. Extreme forms of specific contextual factors help with detecting differences with smaller samples and, when present, provide a means to measure an effect when nonlinearity of confounding variables could dampen experimental findings. Also, extreme forms (or extreme groups) boost statistical power.

One example of an “extreme” physician factor is aging (chapter 3). Does advanced physician age lead to impaired clinical reasoning performance? Importantly, research suggests that physician aging, in and of itself, does not necessarily result in declining clinical reasoning performance. Situated cognition would argue that this is due to the unique and dynamic interplay between this specific physician factor (age) with patient and practice contextual factors. Additionally, studies from fields outside of medicine suggest that crystallized intelligence (the ability to apply contextually appropriate problem solving methods) can increase with aging while fluid intelligence (mental efficiency) declines with advanced age; the result on performance is due to how these changes interact with specific patient and practice factors. Furthermore, studies suggest that older adults are more likely to use adaptive, deep processing strategies and motives than younger adults in their self-regulated learning activities.¹⁵⁻¹⁶ These two important findings—changes in intelligence and typical approach to self-regulated learning activities, combined with their interactions with unique patient and practice factors—suggest that continued professional development must be flexible to changes that are expected to occur during the duration of a physician’s practice.

To help collect additional evidence for our theoretical perspectives and predictions regarding the potential influence of contextual factors on clinical reasoning, two separate investigations were conducted (chapters 4, 6, and 7). As we also sought to develop a form for assessing clinical reasoning to be used in future studies in this thesis, we assessed the feasibility, reliability, and validity of a post-encounter form (PEF; chapters 4 and 5). To provide substantiating evidence for our theoretical viewpoints, we first explored the perspectives of 3 participant groups (medical students, patients, and expert physicians; chapter 4) in a 3-part Objective Structured Clinical Exam or OSCE (i.e. a single clinical encounter). This 3-part OSCE station incorporated each of the three participant group’s views of a single clinical event. We found that different stakeholder groups had similar (but not identical) views of the same event (*r* values of 0.40, 0.32, and 0.42). Also, with comparing decile performance across these three components, a considerable percentage of our second year medical students (31%, 51 of 170 participants) performed in both the top as well as the bottom tercile score on two or more components of this 3-part OSCE. The results from the

confirmatory factor analysis (CFA) were consistent with our expectation that based on theories of situated cognition and ecological psychology--each of the three viewpoints would be significantly, yet moderately, correlated. Further, this study of participant views (chapter 4) involved a post-encounter form (PEF) for assessing clinical reasoning. In chapter 5, we found this same PEF to have high feasibility, good inter-rater reliability and validity evidence with moderate correlations between scores on this rating instrument and related instruments. Based on this reliability and validity evidence from chapters 4 and 5 to support our PEF's use, a version of this form was used in most of our investigations (chapters 4-9).

The results from the investigation in chapter 4 added validity evidence to our theoretical models. An important implication of this CFA study, which was also supported by other investigations in this thesis (chapters 6 and 7), is that obtaining multiple perspectives of the same clinical encounter likely provides a more valid indication of clinical reasoning performance. We believe that based on findings from this study and other investigations discussed below (chapters 4, 6-7), multiple perspectives to evaluate clinical reasoning and to provide feedback are particularly important given that, in essence, we are currently trying to measure a largely internal mental process (clinical reasoning) using tools that capture observable behavior only. Additional implications of this work (chapters 4, 6-7) are that training raters on goals and objectives with observation could improve consistency in ratings and that all participant groups could benefit (and learn) from being explicitly confronted with the contrasting perspectives of the other participants' in the encounter. This study (chapter 4), and additional investigations noted below (chapters 6-7), do help unravel the problem of context specificity—providing evidence that some of this problem is likely due to precision in measurement. In other words, more viewpoints of the same event may attenuate the observed inconsistency in performance from situation to situation (context specificity)—this is a testable hypothesis for future research.

Chapter 4 explicitly explored *different* stakeholder groups and provided evidence to support our theoretical predictions as well as the use of a form for exploring clinical reasoning performance (PEF). It is plausible and predicted by our theoretical models that this phenomenon—similar but not identical views of the same event—would also occur *within* a participant group and contribute to the vexing problem of context specificity. We therefore embarked on a two part, investigation explicitly exploring the influence of contextual factors in expert practice. We studied experts in internal medicine and modified specific contextual factors (patient, practice, and physician factors) in accordance with situated cognition theory in an attempt to address two potential phenomena: quantify the influence of contextual factors (chapter 6) and describe or better understand this influence of contextual factors (chapter 7). Stated another way, how do contextual factors impact expert practice (qualitative study, chapter 7)?

In the quantitative study (chapter 6), we first demonstrated that the problem of context specificity was a part of our findings. We used a modified version of our previously validated PEF (chapters 4 and 5), and we devised a quantitative scoring rubric to capture cognitive load and applied it to think-aloud protocol transcripts as primary outcome measures. For these investigations (chapters 6 and 7), participants viewed a series of videotapes portraying different chief complaints of 3 common diagnoses in internal medicine (HIV positivity, colorectal cancer screening, and diabetes mellitus) while selected patient, practice,

and physician contextual factors were modified. Following each videotape participants completed the PEF and then watched each videotape a second time while undergoing a formal think-aloud analysis.¹⁷

In chapter 6, the effect of the specific contextual factors in the study appeared to be small to moderate with r values within the 0.2 to 0.4 range. Notably, these correlations were negative meaning that specific contextual factors were associated with lower PEF scores. Two notable findings emerged from this study (chapters 6 and 7), results which also appear consistent with how CLT may contribute to the picture of context specificity:

1. The negative associations between the presence of specific contextual factors and lower PEF performance was more likely to be present when the participant verbalized the contextual factor(s) while performing think-aloud analysis. This implies that contextual factor processing (and attendant cognitive load) may influence this negative association with clinical reasoning performance. This implication regarding CLT was also supported by our qualitative investigation (chapter 7), using the same videotapes and same participants. Participants who verbalized one or more specific contextual factors were more likely to report missed key findings on think-aloud analysis. Based on these findings (chapter 6 and 7), it would appear that cognitive load plays an important role in how contextual factors are associated with diagnostic and therapeutic reasoning performance (i.e. cognitive load may be a driving force, or bidirectional arrow [Figure 1], for explaining how context impacts clinical reasoning).
2. Predicted proxy measures for cognitive load on think-aloud analysis (inappropriate use of semantic qualifiers, superficial problem representation, and pauses during the procedure) were associated with lower clinical reasoning performance, when present (r values -0.5 to -0.3). These associations also have implications for novel ways of measuring cognitive load while performing a think-aloud, which, by its design, precludes stopping the procedure to measure cognitive load in a standard way.^{12,17}

In the qualitative arm of this study (chapter 7), we compared think-aloud themes with themes generated from explicit responses to free-text questions inquiring about how contextual factors may influence diagnostic and therapeutic clinical reasoning performance. Identified themes suggested that the contextual factors of patient (low English proficiency, challenges credentials of the physician) and practice (suggest an incorrect diagnosis or treatment, atypical presentation) can be problematic for clinical reasoning performance. On the other hand, these contextual factors can also assist in diagnostic and therapeutic clinical reasoning when only one factor is voiced. Themes from both sources had good overlap, indicating that experts are somewhat cognizant of the potential influences of context on their reasoning processes. There were important exceptions, however, with the themes of *missed key findings*, the *influence of encounter setting*, and *balancing goals*. These themes appeared in the think-aloud analysis but not in the analysis of free-text queries. Notable findings with implications from this study (chapter 7) for our theoretical frameworks were the following:

1. Missed key findings occurred only when participants verbalized one or more contextual factors in the think-aloud protocols. This would argue that the contextual factor may have to be verbalized (or consciously processed) to have an effect on clinical reasoning performance; in essence, the verbalized factor(s) may be taking up working

memory “space” that limits availability of working memory for identifying key findings. This finding also provides additional support for the notion that cognitive load may have influenced the results from quantitative study discussed above (chapter 6). In sum, these two studies suggest that cognitive load is playing a role in both “how” (chapter 7) and “how much” (chapter 6) of an impact our selected contextual factors have on clinical reasoning performance.

2. We constructed a model consisting of *components*, *mechanisms* and *consequences* categories based on our themes. Further, no contextual factor was viewed as beneficial to patient care when more than one was contextual factor was verbalized on think-aloud analysis. One interpretation of this finding again suggests that cognitive load may be playing a role—two (or more) factors equates to more cognitive processing and greater attendant cognitive load.
3. The theme of *balancing goals between participants*, with participant goals that can be fluid or changing, provided important supporting data to the prediction from ecological psychology that participant goals in the clinical encounter can be multiple and can evolve. This theme, as well as the theme of interactions between contextual factors, which would be predicted from situated cognition, provide *mechanisms* by which context might influence clinical reasoning performance and provides another important nuance into the problem of context specificity.

In sum, these findings (chapters 6 and 7) provide important insights and theoretical implications into how and to what degree contextual factors may impact experts’ clinical reasoning performance. These findings were consistent with predictions according to our three main theoretical frameworks. The findings from these studies provide notable insight into the problem of context specificity. More specifically, these results suggest that a component of the problem of context specificity is that contextual factors (patient, practice, physician), other than essential features for establishing the diagnosis, do impact clinical reasoning performance. Further, cognitive load may play an important role in the negative associations found between the verbalized presence of a contextual factor and clinical reasoning performance. Our findings suggest that a second component of the problem of context specificity is that multiple, emerging goals can differ between participants. For example, different participants can have different diagnostic and therapeutic goals, which could lead to differences in clinical reasoning performance and thus contribute to the vexing problem of context specificity. These findings also suggest that cognitive load could be playing a role in other fields for which identifying underlying mechanisms for discrepant findings in medical experts could move the research agenda forward, such as poor self-assessment in physicians^{18,19} despite educational studies which suggest that self-assessment, as a component of self-regulation, is a teachable skill.²⁰ This represents one avenue for future research. With demonstrating important contextual factor effects on expert physician clinical reasoning performance, we then turned to another stakeholder group, medical students, and sought to determine if context of instruction (authenticity of instructional format) impacts clinical reasoning performance.

Research Question 2: Does instructional format impact expertise development in clinical reasoning?

This research question was addressed by a two-year study that explored the balance between increasing the authenticity of the instructional format with negatively impacting cognitive load. Emotional engagement theories would predict that increased authenticity would improve clinical reasoning outcomes, while cognitive load would predict that increased authenticity would increase extraneous cognitive load by adding nonessential verbal and nonverbal information conveyed by the patient. With these competing theoretical predictions, the effect of instructional format (instructional context) on clinical reasoning outcomes was investigated. We followed a cohort of medical students over 2 years.

Increasing authenticity is a patient factor and arguably an *interaction* factor (Figure 1). Cognitive load is a physician factor and arguably an *interaction* factor (Figure 1). In essence, we looked at the problem of context specificity from the standpoint of instruction—does the instructional format matter, leading to variability in student performance across situations of instruction. Another way to view these research questions from the problem of context specificity is to consider the notion that clinical reasoning outcomes following instruction can be seen as a proxy for experience. Prior experience with contextual factors may play a role in the vexing problem of context specificity.

The design was a two-year, prospective, randomized, crossover trial. In chapter 8, we assessed the impact of instructional format on clinical reasoning outcomes at the end of the second year, and in chapter 9 we assessed clinical reasoning outcomes at the end of the third year of medical school. Authenticity of instructional format (paper case, DVD case, or standardized patient portrayal of the same content for three subject areas—abdominal pain, anemia, polyuria) did *not* significantly influence clinical reasoning outcomes at the end of the second year of medical school. These findings suggest that the balance between increasing authenticity of instruction and minimizing cognitive load may be more complex in the preclinical setting than initially conceived though it is also possible that our proxy measures for cognitive load lacked sufficient precision, as we did not incorporate standard measures of cognitive load because the importance of CLT to this research question only became evident as the study progressed.

Authenticity, as well as cognitive load, may vary to a great degree on a person-by-person basis, and the specifics of the environment. Studies exploring the underpinnings of this finding, to include the use of validated measures of cognitive load and emotional engagement instruments, could help illuminate this balance more fully. An important implication of this finding is that despite the marketing by the simulation industry, and our desire to reach an increasingly technology-oriented trainee population, increasing instructional authenticity is not a panacea.

Exploring the third-year outcomes (chapter 9) revealed that second-year instructional format did not impact third-year clinical reasoning outcome measures. However, the number of patients seen during the third-year clerkship did impact third-year clinical reasoning outcomes. Notably, there was no crossover effect on our outcome measures—seeing more patients with anemia was not associated with improved performance for diabetes or abdominal pain outcomes (i.e. context specificity).

Combining the preclinical and clinical results (chapters 8 and 9, respectively), one conclusion is that we found additional evidence for the need for deliberate practice to be successful on clinical reasoning outcomes. This has been shown in numerous studies regarding expertise from fields outside of medicine.¹⁷ Deliberate practice argues that to obtain expertise, effortful practice of components of the activity must be performed, and, at least at first, this is done through the aid of a coach or mentor.¹⁷ Another conclusion from these investigations (chapters 8-9) is the need for deliberate practice with authentic instructional formats, which is consistent with our notion that contextual factors matter—contextual factors must be “practiced” to be “mastered” (as measured by clinical reasoning outcomes). This finding is also consistent with the tenet of high-element interactivity that would be predicted by our overarching theories with learning in the clinical environment. Moreover, number of patients recently seen with the diagnosis portrayed in the expert videotape study (chapter 6) did not influence participants’ clinical reasoning performance. In line with the assumptions of CLT, perhaps seeing additional patients led to no further decrement in element interactivity (and thus reducing cognitive load) or additional automation (also reduces cognitive load), whereas for third-year students this reduction in element interactivity and/or additional automation through deliberate practice was crucial to performance by reducing cognitive load.

Notably, the clinical reasoning literature does not provide guidance on instructional practices that address the various contextual factors inherent to the clinical encounter. We believe faculty development in this area could improve the acquisition of clinical reasoning expertise, as it would provide a means for the faculty “coach” or mentor to guide and instruct the trainee, focusing their (practice) learning on the nuances of contextual aspects of clinical practice. The emerging Accreditation Council of Graduate Medical Education (ACGME) competencies of practice-based learning and improvement and systems-based practice do, in some respects, get at the need for training and reflection on contextual factors in the clinical encounter.²¹ We believe these context of instruction findings (chapters 8 and 9) add another piece to the puzzle of context specificity. Based on our findings, we would hypothesize that prior instructional experience (relative expertise) with the constellation of contextual factors in a specific situation (a specific clinical encounter) may influence performance, as much as or even more than the specific diagnosis being portrayed.

It is plausible that problem representation and formulating a differential diagnosis, at times, are highly complex and potentially nonlinear processes—this would be consistent with situated cognition (multiple emerging views), ecological psychology (multiple and changing goals) and CLT (ability to manage emerging data and element interactivity in working memory as well as extraneous load). These processes may entail more than seeking for and then confirming a diagnostic script in the physician’s head as predicted by information processing theories.²² The ability to attend to, process, interpret and manage information in the clinical encounter with current information processing theory models implies a stability in the individual that is independent of the situation; our findings (chapters 2-9) would suggest that this may not be the case.

Research Question 3: Can we improve upon assessment of clinical reasoning by triangulating behavioral measurements with concurrent verbal reports (think-aloud protocols) or with novel instruments such as fMRI scanning?

We used a think-aloud protocol in several investigations (chapters 6-7, and 10) to further enlighten our understanding of context on a complex mental process (clinical reasoning). In one study (chapter 10), we combined findings from functional magnetic imaging (fMRI) and a think-aloud protocol to explore clinical reasoning expertise. Participants were board-certified internists who answered live licensing exam questions in an fMRI scanner and then went through a formal think-aloud protocol reviewing the items answered in the scanner.

CLT proposes that schema automation is a key means to reduce cognitive load.²³ Expertise is believed to be associated with rich, automated schemas.¹⁷ In clinical reasoning theory, non-analytic reasoning refers to automated reasoning which is believed to play a prominent role in the expert. Non-automated reasoning, also known as analytic reasoning, is effortful and involves the deliberate comparing of different options; it is believed to play a prominent role in the novice. We therefore sought to determine if there are functional neuroimaging differences consistent with expertise.

There were several important findings from the fMRI study (chapter 10) to inform future work. First, we found limited evidence to support theories of analytic and non-analytic reasoning, and patterns of neuroimaging activation were consistent with limited work outside of medicine. Specific fMRI findings revealed significantly increased activation of prefrontal cortex with answering incorrectly versus correctly, activation of precuneus and mid-temporal gyrus with not guessing versus guessing, as well as some differences between answering versus reflecting upon answers. Second, our results suggest that for clinical reasoning, experts may use multiple areas of working memory, located in several parts of the brain, to include the cerebellum. This is consistent with literature on working memory^{24,25} as well as CLT's principle of multi-modal learning.^{12,13} Third, our results suggest that expertise in internal medicine does not have a uniform functional neuroimaging pattern. Fourth, the brain processes corresponding to established theories may be more complex and subject to wide inter-individual differences--it is therefore possible that the differences may be nonlinear. Nonlinearity in mental processes has been seen in other fMRI studies.²⁶⁻²⁹ Finally, the finding of few differences between answering and reflecting on answers adds further validity evidence to the use of a think-aloud protocol for understanding cognitive processes like clinical reasoning; the perceived bias in the effects on thinking induced by asking a participant to think-aloud may not be present for the task of clinical reasoning.¹⁷

In the first think-aloud study (chapter 6) we used a deductive coding scheme developed a priori, geared at identifying verbalization of the presence of one or more contextual factors under study in addition to other potential measures of cognitive load, so that we could compare these think-aloud findings with other quantitative measures. In the second, qualitative, study (chapter 7) the think-aloud data were analyzed in a constant comparative, inductive fashion.³⁰ Findings from the qualitative study (chapter 7) suggested that increased cognitive load may have lead to the theme of key missed findings. We therefore included potential markers of cognitive load for our quantitative study (chapter 6) so that we could further explore this phenomenon of key missed data from chapter 7 and because think-

aloud protocols preclude the intermittent use of standard measures of cognitive load, which would detrimentally interrupt the think-aloud procedure and negate findings. In both chapters (chapters 6 and 7), the think-aloud data provided additional evidence that corroborated our other measures and also provided data that added to the other measures under investigation. More specifically, in addition to fMRI data providing validity evidence, we found preliminary validity data to support the use of the think-aloud protocol for measuring cognitive load (chapter 6) that should be confirmed in future studies.

These findings (chapters 6, 7 and 10) provide additional data to help unravel the problem of context specificity. Not only do these findings concur with our CFA addressed in chapter 4 suggesting the need for multiple sources of data to deal with the issue of context specificity, but these findings (chapters 6, 7 and 10) also suggest that think-aloud and fMRI analysis, in particular, could be incorporated into the assessment of clinical reasoning to help address measurement precision, which we believe is a part of the problem of context specificity. Though our findings are encouraging for future research, until the cost of fMRI is reduced or if subsequent findings suggest that specific participant groups would be likely to benefit from this approach (i.e. those who struggle academically), this combined methodology using fMRI does not appear to be feasible for educational practice. In other words, we believe that for fMRI, the answer to this research question is “yes” for research purposes but “no” for educational practice (at least at the current time).

Conclusions

From our investigations, several conclusions can be drawn. In addition to the conclusions associated with the above discussion of each research question, to include how results from specific investigations enlighten the problem of context specificity, we also propose the following overarching conclusions:

1. Context is more than a static physical location or additional information surrounding the presenting complaint. Context in the clinical encounter is a complex system that involves both social (participants) and physical (setting) constructs, all of which evolve over time. This thesis provides some additional theoretical understanding as well as empirical evidence to support this viewpoint. Both potential reasons why context impacts clinical reasoning performance (chapter 7) as well as some understanding of magnitude of the effect (chapter 6) are discussed. As these investigations involved experts, we do not know how these contextual factors may influence residents and even medical students. We would predict from our theoretical model and our data that the effects on performance would be greater, but this hypothesis requires further exploration.
2. We need to consider the effects of instructional context on clinical reasoning performance—the above-stated balance between instructional authenticity and cognitive load is relevant. Such an effect is predicted from our three theoretical frameworks. Further, our work suggests that this balance appears to be dependent on the relative expertise level of the participant, as we found different results with different participant group levels (second versus third-year students). Cognitive load theory has also found different results for the same task dependent upon the expertise level of the participant, the

so called “expertise reversal effect”.¹³ We believe this adds further validity evidence to our findings. Therefore, the teaching context matters for the development of expertise in clinical reasoning, and one-size-fits-all instructional approaches are of limited effectiveness given that both authenticity and cognitive load are likely more personal than participant group dependent phenomena.

3. Our practice and teaching settings may display nonlinearity at times and potentially even deterministic chaos. This conclusion is consistent with our overarching theoretical frameworks, which would predict nonlinearity at times as the outcome is driven by interactions and feedback among factors in the environment (patient, physician, practice). These interactions and feedback are not predictable at the outset, emerge as the encounter evolves, and therefore can be nonlinear. Our empirical findings in the studies presented here also do not rule out the possibility that such nonlinearity exists.
4. Combining measurement strategies appears to be a helpful way to improve our understanding of clinical reasoning, a complex construct. We would also encourage the consideration of other techniques from fields outside of medical education to help improve our understanding of the largely internal mental process of clinical reasoning for research purposes. Examples could include, but are not limited to, physiologic measures (respiratory rate, galvanic skin response, blinking rate) and facial feature analysis (Facial Action Coding System or FACS). These measures could provide additional clues into internal mental processes like clinical reasoning.
5. Established theoretical frameworks are needed to improve our understanding of complex human phenomenon such as clinical reasoning. We believe that by combining situated cognition, ecological psychology, and CLT, we have provided a potentially useful model for designing, analyzing, and predicting how contextual factors impact clinical reasoning.
6. Though not an emphasis of our study, our data would suggest that diagnostic and therapeutic reasoning may be different constructs. Two investigations (chapters 6 and 7) indicated that contextual factors appear to have less of an influence on therapeutic than on diagnostic reasoning.
7. Expertise appears to be a state (i.e. it depends on the specific *situation*) as opposed to a static trait. As such, we would advocate for the term *expert performance* as opposed to *expertise* to convey the state nature of this phenomenon. In our studies of experts, an individual’s performance across items and cases (chapter 10, and chapters 6 and 7, respectively) varied greatly. This was helpful for our studies, as we assessed participants on both an inter-individual and intra-individual basis, leading to less range restriction with interpreting results. For the study involving videotapes, it is conceivable that contextual factors contributed to this phenomenon. In other words, deliberate practice in both the various manifestations of disease presentation as well as contextual factors may be needed for expertise development and even expertise maintenance. This would represent another area for future investigation.

Limitations and recommendations for future research

There were several important limitations to our work:

1. For the majority of investigations, we used relatively small samples consisting of a single participant group (i.e. expert internists or medical students). Future work could strive to include other participant groups to see if similar results emerge. For example, including residents in the fMRI study (chapter 10) or residents and even medical students in the contextual factors study (chapters 6 and 7). Also, studies using the same or similar design could be repeated with a larger sample to ascertain if additional findings emerge.
2. Though cognitive load was a theoretical framework used in several studies (chapters 6-9), we did not include formal measures of cognitive load.^{12, 13, 23} For studies incorporating think-aloud protocols, we did not think a standard measure of cognitive load would be appropriate as it could compromise results. However, future work directly comparing our think-aloud coding scheme measures with standard instruments for measuring cognitive load could inform both procedures.
3. We raised the important issue in this thesis regarding the possibility, at times, of nonlinearity or even deterministic chaos in the clinical encounter. Future work could experimentally seek to gather evidence to confirm or refute this possibility.
4. We studied a small sample of contextual factors proposed by the situated cognition framework. Certainly future work could build upon our findings by using the same or similar study designs with additional contextual factors to see if similar findings are observed.
5. Our studies occurred in somewhat controlled settings. We believe that this was necessary to provide proof of concept for our theoretical frameworks, determine if expected predictions occurred, and to develop and gather reliability and validity evidence for our measurements. Given our findings and theoretical frameworks, future studies could be designed to measure the impact of contextual factors in actual clinical practice.
6. Additional measurements from fields outside of medical education could provide new understanding to our line of research. These could include facial feature analysis and obtaining physiologic measurements. This additional information could provide important triangulating data to help us measure something that we cannot physically observe—the mental processes involved with clinical reasoning.
7. The focus of this thesis was on diagnostic reasoning and our findings suggest diagnostic and therapeutic reasoning may not be the same. Future investigations could specifically explore the influence of contextual factors on therapeutic reasoning.
8. Our work primarily focused on students and experts from a single institution. Results may not be generalizable to other institutions.
9. Though we did find reliability and validity evidence to support the use of the post-encounter form used in several chapters (4-9), the evidence was modest.
10. Finally, our thesis, was more “light bearing” (to improve understanding) than “fruit bearing” (explore interventions based on understanding).³¹ We believe this was warranted, as theoretical frameworks and empirical evidence to support our theoretical stances was needed prior to investigating, for example, whether or not a designed in-

tervention improves learning and performance. This would be an important next step as we seek to better understand and investigate context specificity.

Implications

Our theoretical frameworks and study findings provide some insights into the influence of contextual factors on clinical reasoning and context specificity. From these findings, we would suggest the following additional implications.

1. One way to better understand the problem of context specificity is to challenge the assumption of stability in the individual and stability in the environment. Indeed, our frameworks and findings would suggest the contrary. We would propose that it is persons, environment, and their interactions that matters with making diagnostic and therapeutic decisions (Figure 1). We believe that future studies and educational practices to test and inform these assumptions would benefit from further investigation of the person plus the environment.
2. Nonlinear, evolving, complex interactions may occur during the clinical encounter. Empirical evidence to support or refute this phenomenon is needed. The concept of nonlinearity may not only shape our discourse in practice but also how we frame problems and potential solutions. It may be that we are attempting to make something linear, when it is not. On this point, it could be that experts not only attend to more information in the environment (when compared to novices), but maybe they also have the ability (presumably learned) to recognize a key piece of information in a nonlinear system, that will ultimately impact their diagnostic and therapeutic decisions. Indeed, the “ah ha” moment in reaching a correct diagnosis and therapy may, at times, be a nonlinear event. Regardless, considering the environment and its interactions with participants, means that the true expert is flexible, taking into account not only the clinical presentation but also recognizing the unique contribution of interactions—it is not either or: success requires an understanding of the unique contributions of clinical presentation and complex interactions. An important implication of our notion of context and the clinical encounter is the need for real time feedback from multiple sources to allow the “physician factor” to “correct” or learn in this complex system.
3. In several chapters, we collected evidence to support the role of CLT in clinical reasoning—both in terms of the impact of contextual factors on experts’ clinical reasoning performance (chapters 6 and 7), as well as the impact of context on developing clinical reasoning expertise (chapters 8 and 9). As such, we believe that efforts to reduce cognitive load, in both clinical practice and teaching settings, is paramount. There are a number of recent works that describe various teaching strategies for reducing cognitive load.^{12, 13, 23}
4. Faculty development efforts are needed—not only to teach faculty about the potential impact of contextual factors on clinical reasoning for their own practice but also to train future generations of physicians. This training should not only involve instruction on how contextual factors can impact reasoning, but it should also help both faculty and trainees recognize the presence of contextual factors that may influence reasoning from our work. The context of clinical reasoning instruction also matters—increasing authenticity should not be assumed to lead to improved performance. For example,

having students see actual patients the first year of medical school, an instructional approach that is being contemplated by many US medical schools, may not be effective. Further, understanding that different participants in the encounter can have different views of the same event (chapter 4) provides important implications regarding feedback. Not only do our findings suggest that feedback needs to be timely and originate from multiple sources, they also suggest that the teacher needs to pay particular attention to ensuring understanding and repeating key learning points in multiple ways as the learner may have difficulty translating feedback from one context to the next.

5. Traditional, passive approaches to continuing medical education have been ineffective.¹⁹ It should follow that we need to use our current understanding of physician factors such as aging, cognition, and the emerging expertise literature to affect changes in both practice and lifelong learning across the continuum of a physician's career. We believe that situated cognition is an inclusive framework that incorporates practice, patient, and physician factors that interact in dynamic ways.
6. At times, investigators may need to combine existing theories or build new theories to help illuminate challenges in medical education and clinical practice. We believe that our theoretical framework, which combined situated cognition, ecological psychology, and CLT, provides a useful lens for exploring the vexing problem of context's impact on clinical reasoning (context specificity).

Ultimately, we believe that unraveling the problem of context specificity will require better understanding of the *underlying mechanisms* that drive this phenomenon. We believe this is analogous to understanding the pathophysiology of disease. Ground-breaking diagnostic and therapeutic advancements have been made with understanding complicated pathophysiology for diseases such as cancer, unstable angina, and stroke. Furthermore, such investigations led to understanding that some conditions that were assumed to be the same were indeed different. Examples include recognizing that heart failure can be systolic or diastolic. Also, conditions assumed to be different can be very similar--unstable angina and acute myocardial infarction share largely the same pathophysiology. Clinical outcome prediction and treatment is improved with mechanistic understanding in clinical practice. Further, with mechanistic understanding, clinical problems that were assumed to represent one condition have been found to represent two (or more) conditions (i.e. heart failure), and problems that were assumed to represent different conditions were found to represent similar or the same condition (i.e. unstable angina and acute myocardial infarction).

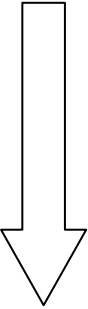
Along these lines, we believe that mechanistic understanding of the complicated problem of context specificity could lead to important advancements in outcome prediction ("diagnosis"; clinical encounter and educational) and "treatment" (physician trainee and expert). Further, additional exploration into interacting factors (bidirectional arrows in Figure 1), may be particularly fruitful. Approaches to the problem of context specificity have largely focused on the below linear model:



What contributes to the arrow? This is at the heart of the issue of context specificity, as the same presenting problem can lead to different solutions in experts and trainees. Situated cognition and ecological psychology propose that the problem is nested in the *situation* (patient, practice, and physician factors) and their interactions. The interactions may be particularly relevant, as they represent potential *mechanisms* or driving forces behind the problem of context specificity in clinical reasoning. Indeed, we have found that cognitive load, authenticity, balancing participant goals, interactions between factors, and the dyad of affordances and effectivities are some of the potential mechanisms underlying this phenomenon. Furthermore, additional investigation of the *components* influencing the impact of contextual factors and *consequences* of these contextual factors (chapter 7) could also lead to important mechanistic insights to help improve our understanding of the impact of contextual factors on clinical reasoning.

For example, with cognitive load, what would be a “treatment” approach given its identification as a likely mechanism underlying context specificity? A series of studies could follow the below sequence, which could both lead to further understanding of the mechanism (“diagnosis”) as well as potentially improve expert and trainee performance (“treatment”).

Figure 2: Authenticity, cognitive load, clinical problem solving complexity and contextual factors interact*



Authenticity	Contextual factor complexity	Problem complexity
Simulated design	single contextual factor	well defined problem
Simulated design	single contextual factor	ill defined problem
Simulated design	multiple contextual factors	well defined problem
Simulated design	multiple contextual factors	ill defined problem
Authentic design	single contextual factor	well defined problem
Authentic design	single contextual factor	ill defined problem
Authentic design	multiple contextual factors	well defined problem
Authentic design	multiple contextual factors	ill defined problem

*=a potential sequence of studies to further understand how authenticity, cognitive load, clinical problem complexity and contextual factor complexity interact

We believe the single trajectory or single correct response as implied by the single, unidirectional above arrow represents a very narrow and arguably naive view. The investigations in this thesis demonstrate that contextual factors *do* impact performance, and that these factors likely do so through a variety of dynamic mechanisms predicted by our empirical data and theoretical frameworks. In this light, we have doubts about two or more physicians following exactly the same linear trajectory even if they encountered the same situation (identical context, see Figure 1). For example, it would be reasonable to expect that two physicians would not have identical goals, affordances and effectivities, and cognitive load in the situation and therefore may not arrive at identical diagnostic and therapeutic plans. This is not unlike asking a group of major league baseball pitchers to throw an identical pitch to an identical location in the strike zone; it is very unlikely to occur.

We would, however, argue that a group of experts seeing the same patient in an identical situation (identical context, Figure 1) would likely have *similar* trajectories (i.e. most arrive at reasonable diagnostic and treatment plans), just like a group of major league baseball pitchers would likely throw the ball to a very similar, but not identical, location—most would still pitch a “strike”. Given our proposed theoretical frameworks and empirical findings, we would argue that the optimal clinical outcome for an encounter is not predictable at the outset of the appointment but it is a “bounded” condition (there is a “strike zone”). On this point, we believe part of the problem of context specificity is the fact that context cannot be completely controlled, reduced, or unraveled because of the complexity of the clinical encounter that has been discussed in this thesis. We do believe, however, that investigating this phenomenon can lead to important insights to help mitigate unwarranted variance (i.e. key missed findings) that could ultimately improve patient care. In other words, we want our expert physicians to consistently “throw strikes” and show our physicians in training how to most efficiently learn how to do so, because the practice of medicine is too important to “lose”—we have a societal obligation to better understand the problem of context specificity so that our patients receive the care that they need and deserve.

References

1. Lave L, Wenger E. *Situated Learning: Legitimate Peripheral Participation*. New York (NY): Cambridge University Press; 1991.
2. Brown JS, Collins A, Duguid P. Situated cognition and the culture of learning. *Educational Researcher*. 1989;18:32-42.
3. Bredo E. Reconstructing educational psychology: situated cognition and Deweyian pragmatism. *Educational Psychol*. 1994;29:23-35.
4. Robbins P, Aydede M. *The Cambridge Handbook of Situated Cognition*. New York (NY): Cambridge University Press; 2009.
5. Heft H. *Ecological Psychology in Context: James Gibson, Roger Barker, and the Legacy of William James' Radical Empiricism*. Mahwah (NJ): Erlbaum; 2001.
6. Young MF. Instructional design for situated learning. *Educ Technol Res Dev*. 1993;41:43-58.
7. Jonassen D, Land S, eds. *Theoretical Foundations of Learning Environments*. Mahwah (NJ): Erlbaum; 1999.
8. Gleick J. *Chaos: Making a New Science*. New York (NY): Viking Penguin Inc; 1987.
9. Thelen E. Development as a dynamic system. *Current Directory of Psychol Sci*. 1992;1:189-93.
10. Eve RA, Horsfall S, Lee ME, eds. *Chaos, Complexity, and Sociology: Myths, Models and Theories*. Thousand Oaks (CA): Sage Publications; 1997.
11. Paas F, Renkl A, Sweller J. Cognitive load theory and instructional design: recent developments. *Educ Psychol*. 2003;38:1-4.
12. van Merriënboer JJG, Sweller J. Cognitive load theory and complex learning: recent developments and future directions. *Educ Psychol Rev*. 2005;17:147-77.
13. van Merriënboer JJG, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ*. 2010;44:85-93.
14. Personal communication with Mike Young, 2010.
15. Richardson JC, Newby T. The role of student's cognitive engagement in online learning. *Am J Dist Educ*. 2006;20:23-37.
16. Greene JA, Azevedo R. A theoretical review of Winne and Hadwin's model of self-regulated learning: new perspectives and directions. *Rev Educ Res*. 2007;77:334-72.
17. Ericsson KA, Charness N, Feltovich P, Hoffman RR, eds. *The Cambridge Handbook of Expertise and Expert Performance*. New York (NY): Cambridge University Press; 2006.
18. Eva KW, Regehr G. Self-assessment in the health professions: A re-formulation and research agenda. *Acad Med*. 2005;10 Suppl:S46-54.
19. Davis DA, Mazmanian PE, Fordis M, Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*. 2006;296:1094-102.
20. Zimmerman BJ. Attaining self-regulation: a social-cognitive perspective. In: Boekaerts M, Pintrich P, Zeidner M, eds. *Handbook of Self-regulation*. Orlando (FL): Academic Press; 2000:13-39.
21. ACGME.org [Internet]. [cited 2011 Jan].
22. Durning SJ, Artino AR. Situativity theory: a perspective on how participants and the environment can interact. *AMEE Guide 52. Med Teach*. 2011;33:188-199.
23. Plass JL, Moreno R, Brunken R, eds. *Cognitive Load Theory*. New York (NY): Cambridge University Press; 2010.
24. Baddeley AD. *Working Memory*. Oxford: Clarendon Press; 1986.
25. Baddeley A. Working memory. *Science*. 1992;255:556-9.
26. Friston KJ, Mechelli A, Turner R, Price CJ. Nonlinear responses in fMRI: the balloon model, volterra kernels, and other hemodynamics. *NeuroImage*. 2002;12:466-77.
27. Janz C, Heinrich SP, Kornmayer J, Bach M, Hennig J. Coupling of neural activity and BOLD fMRI response: new insights by combination of fMRI and VEP experiments in transition from single events to continuous stimulation. *Magnet Reson Med*. 2001;46:482-6.
28. Kelso JAS. *Dynamic Patterns: The Self-organization of Brain and Behavior*. Cambridge (MA): MIT Press; 1995.
29. Lauritzen M, Gold L. Brain function and neurophysiological correlates of signals used in functional neuroimaging. *J Neurosci*. 2003;23:3972-80.
30. Kennedy TJT, Lingard LA. Making sense of grounded theory in medical education. *Med Educ*. 2006;40:101-8.
31. Zittel C, Engel G, Nanni R, Karafyllis NC. *Philosophies of Technology: Francis Bacon and His Contemporaries*. Leiden (Neth): Koninklijke Brill NV; 2008.

Chapter 12

Summary

The purpose of this thesis was to explore the influence of contextual factors on clinical reasoning during the clinical encounter. More specifically, our overarching research question was: do selected contextual factors impact clinical reasoning performance, both in expert practice and during instruction of trainees?

Chapter 1 outlines the main research question and also introduces the specific research questions that are addressed in this thesis. A primary focus of this thesis was unraveling the vexing problem of context specificity as it pertains to clinical reasoning performance. This unraveling includes exploring the problem of context specificity germane to expert performance, instructional context, and methods of assessment. We first discuss what is known about the reliable and valid measurement of clinical reasoning, and we introduce the problem of context specificity, which challenges our measurement of clinical reasoning in clinical practice. Next, context is discussed in more detail and two overarching theoretical frameworks are introduced, situated cognition and ecological psychology, which serve as a main through-line to the investigations conducted in this thesis. From these two theoretical viewpoints, it is assumed that one cannot fully appreciate clinical reasoning without considering the context, since reasoning is located in experience (participants, practice, and their interactions). Importantly, we did not attempt to measure reasoning from a standard psychometric sense in this thesis—our emphasis was on a better understanding of this pervasive phenomenon in medical education. Thus, this thesis explored the influence of context on clinical reasoning performance. In this light, we emphasized the importance of theoretical frameworks to provide a lens for future investigation.

Chapter 2 begins with a challenge that physicians' face--the idea that training and practice occur in complex environments. We discuss how studies from fields outside medicine can be used to approach the notion of context in the clinical encounter. First, we explore the use of the term context in the clinical encounter as it relates to medical education. Next, we detail the meaning and use of the term context in diverse fields outside medicine, such as mathematics, physics, and psychology, and we provide a definition for context that pertains to medical education and which is used to frame all the studies in the present thesis. We then introduce two inclusive theories, situated cognition and ecological psychology, for the purpose of studying context in medical education. The theories illustrate how individual components of the clinical encounter, or contextual factors (patient, physician, and practice), can relate to the outcome (patient care) in potentially complex ways. Considering contextual factors from this perspective--patient, physician, practice, and their interactions--is consistently applied in this thesis. Next, we highlight several of these factors and their interactions in the clinical arena. We highlight challenges and implications in the field of medical education. An important point is that these theories argue for the need to consider nonlinear approaches when exploring context, where the outcome can equal more than the sum of the individual parts (contextual factors). Finally, we propose next steps for research regarding context for this thesis and medical education in general.

Chapter 3 addresses an example of a physician contextual factor—aging. We explore this factor from the standpoint of the situated cognition framework, with the purpose of understanding potential challenges and implications of our theoretical model on aging physi-

cians. More specifically, we outline several issues related to clinical reasoning performance and implications associated with aging physicians. This includes outlining the potential impact of aging on expertise, information processing, neurobiology, intelligence, and self-regulated learning. We believe that work done in these related fields can provide a better understanding of physician aging and clinical reasoning performance, and thus can inform more effective approaches to continuous professional development and lifelong learning in medicine. We conclude with implications for the health care system and areas of future research.

Chapter 4 reports on a quantitative study that explored three participant views of the same event (patient, physician, and trainee) to determine if these different stakeholder groups view the same event differently. Situated cognition and ecological psychology would propose that these three stakeholder groups would have different, yet somewhat related viewpoints, as these three groups have different affordances and effectivities (ecological psychology) and would be expected to interact in unique ways with the patient, encounter setting, and each other (situated cognition). We conducted a confirmatory factor analysis (CFA) of a three-component OSCE station to ascertain patient, physician, and trainee viewpoints of the same event. Consistent with our theoretical frameworks, we observed moderate correlations between the three views. Also consistent with our theories, individual trainee performance, as measured by tercile score, markedly varied across components of the OSCE station. We conclude that obtaining multiple perspectives of the same encounter may provide a more valid measure of a student's clinical performance, a conclusion that is supported by our theoretical frameworks and study findings.

Chapter 5 is a quantitative study that provides evidence to support the feasibility, reliability, and validity of a post-encounter form (PEF) to assess clinical reasoning. This PEF was also used in chapter 4 and in subsequent chapters. This instrument development paper explored psychometric evidence for the free-text PEF in an Objective Structured Clinical Examination (OSCE) for end-of-second-year medical students. Feasibility was high, as assessed by time needed to grade the form. Inter-rater reliability was high, as assessed by kappa. Validity evidence was obtained by performing univariate comparisons between the individual items on the PEF and other components in the 3-part OSCE station (i.e., standardized patient checklist and oral presentation rating form). Additional validity evidence was found when comparing scores on the PEF with other end of second year grades. We concluded that our PEF yielded adequate psychometric properties and, therefore could be used in the evaluation of clinical reasoning. As such, the PEF was used in subsequent chapters of this thesis.

Chapter 6 reports on a mixed-methods design investigation that explored the impact of selected patient, physician and practice or encounter contextual factors (low English proficiency, challenging the credentials of the physician, incorrect diagnostic or therapeutic suggestion and atypical presentation of the diagnosis). In the quantitative study, (chapter 7 represents the qualitative portion of this study) we first demonstrated that the problem of context specificity was observed in our findings. We used a modified version of the above PEF (chapters 4 and 5) as an outcome measurement. We also devised a quantitative scoring

rubric and applied it to think-aloud protocol transcripts. For both components of this study (chapters 6 and 7), experts in internal medicine (board-certified internists) viewed a series of videotapes portraying different chief complaints of three common diagnoses in internal medicine (HIV positivity, colorectal cancer screening, and diabetes mellitus) while selected patient, encounter, and physician contextual factors were modified. Following each videotape, participants completed the PEF and then watched each videotape a second time while undergoing a formal 'think-aloud' analysis. The effect of the specific contextual factors appeared to be small to moderate, with r values ranging from 0.2 to 0.4. Two important findings emerged from this study; these findings are consistent with how cognitive load theory (CLT) may contribute to the problem of context specificity. First, the negative correlations between the presence of contextual factors and PEF performance was more likely to be seen when participants verbalized contextual factor(s) during think-aloud analysis. Second, predicted surrogate measures for cognitive load on think-aloud analysis (inappropriate use of semantic qualifiers, superficial problem representation, and pauses) were associated with lower clinical reasoning performance as measured by our PEF.

Chapter 7 is the qualitative component of the investigation conducted in chapter 6. The purpose of this study was to explore how contextual factors may influence clinical reasoning. Qualitative data about experts were gathered from two sources: think-aloud protocols and free-text queries that explicitly asked our board-certified internists to comment on the influence of selected contextual factors on their clinical reasoning processes. These data sources provided both actual performance data (think-aloud responses) and opinions on reflection (free-text answers) regarding the influence of context on reasoning. Results for each data source were analyzed for emergent themes using a constant comparative approach and then combined into a unified theoretical model. Several themes emerged and were broadly classified as *components* influencing the impact of contextual factors, *mechanisms* for addressing contextual factors, and *consequences* of contextual factors on patient care. Themes from both data sources had good overlap, suggesting that experts are at least somewhat cognizant of the potential influences of context on their reasoning processes. Notable exceptions were the themes of *missed key findings*, *balancing goals*, and the *influence of encounter setting*, which appeared in the think-aloud analysis but not in the free-text queries. Missed key findings occurred only when participants verbalized one or more contextual factors in the think-aloud protocols, suggesting that cognitive load may play a role. Our unified model is consistent with tenets of cognitive load, situated cognition, and ecological psychology theories and a number of potentially modifiable influences of contextual factors on clinical reasoning are discussed. The results from both chapters (6 and 7) suggest that a component of the problem of context specificity is that contextual factors (patient, practice, physician), other than key features for establishing the diagnosis, do impact clinical reasoning performance.

Chapter 8 reports on part one of a two-part (chapters 8 and 9) prospective randomized crossover trial exploring the impact of authenticity of teaching format (teaching context) on clinical reasoning outcomes in second-year medical students. We explored the balance between increasing authenticity and cognitive load. In essence, we viewed the problem of context specificity from the standpoint of instruction—does the instructional format mat-

ter, leading to variability in student performance across situations of instruction. Increasing authenticity of instructional format (paper case, DVD case, or standardized patient portrayal of the same content) did not influence outcome measures for clinical reasoning, with none of the outcome measures reaching statistical significance. These findings suggest that the balance between increasing authenticity of instruction and minimizing cognitive load may be more complex in the preclinical setting than initially conceived. Medical educators should therefore consider balancing increasing authenticity with factors like cognitive load and learner experience when designing new instructional formats. Another important implication of these findings is that despite intense marketing by the simulation industry, and our desire to reach an increasingly technology-oriented trainee population, increasing instructional authenticity is not a panacea.

Chapter 9 explores the same cohort of medical students in chapter 8 through their third-year of medical school. Second-year instructional format (chapter 8) did not impact third-year clinical reasoning outcome measures. However, the number of patients seen during the third-year clerkship did significantly impact third-year clinical reasoning outcomes. Importantly, there were no crossover effects on our outcome measures—seeing more patients with anemia was not associated with improved performance for diabetes or abdominal pain outcomes. This chapter adds to the literature by demonstrating that authenticity of preclinical teaching format does not appear to matter for the clerkship year. However, actual patients seen (what might be considered the most authentic format of all) does appear to influence clinical reasoning clerkship outcomes.

Chapter 10 looks at the challenge of context specificity in experts using a novel measurement approach, functional magnetic resonance imaging (fMRI). In essence, part of the context specificity problem in clinical reasoning may be that our measures have insufficient precision to capture these largely internal mental processes. In this chapter, we combine fMRI and think-aloud procedures to determine if there were functional neuroimaging correlates to analytic and non-analytic reasoning theory, which is consistent with the concept of automation from cognitive load theory, and to determine if there are functional neuroimaging differences with answering versus thinking aloud on vignette-based test items. Board-certified internists answered live licensing exam questions in an fMRI scanner followed by completion of a formal think-aloud procedure. Evidence, in the form of fMRI changes, to support the presence of analytic and non-analytic reasoning systems was found. We also found evidence to support the use of think-aloud protocols to capture internal mental processes such as clinical reasoning. This model of employing functional neuroimaging activation correlates may enhance our understanding of theory as well as our practice of teaching and assessing clinical reasoning.

Chapter 11 explains how all previous chapters contribute to exploring the influence of contextual factors on clinical reasoning during the clinical encounter. It returns to our overarching question, which was do contextual factors impact clinical reasoning performance, both in expert practice and during instruction of trainees? The answer from our work is that contextual factors do impact expert and learner clinical reasoning performance. Context in the clinical encounter is a complex system that involves both social (parti-

pants) and physical (setting) constructs, all of which can evolve over time. This thesis provides some additional theoretical understanding, as well as empiric evidence, to support this phenomenon. A main focus of this thesis was improving our understanding of the problem context specificity as it pertains to clinical reasoning performance, and several pieces to this puzzle are discussed. This understanding includes exploring the problem of context specificity germane to expert performance, instructional context, and methods of assessment. It subsequently discusses the two overarching theoretical frameworks, which contributed to our understanding of contextual factors and how cognitive load theory emerged as an important third framework for our investigations. Conclusions are then discussed, as well as implications, limitations and recommendations for future research. Overall, our findings were consistent with predictions according to our three main theoretical frameworks, and a number of future areas of exploration are discussed.

Samenvatting (Summary in Dutch)

Het doel van dit proefschrift is te onderzoeken wat de invloed is van contextuele factoren op het klinisch redeneren tijdens patiëntenconsulten. Specifieker gesteld, was onze onderzoeksvraag: "Hebben bepaalde contextuele factoren een invloed op het vermogen klinische te redeneren, zowel in de dagelijkse specialistenpraktijk als bij het onderwijs aan studenten en assistenten?"

Hoofdstuk 1 geeft een overzicht van de centrale onderzoeksvraag en leidt de specifieke onderzoeksvragen in, zoals die in dit proefschrift aan de orde komen en onderzocht worden. Het primaire doel van deze dissertatie was het ontrafelen van het lastige probleem van contextspecificiteit bij klinisch redeneren. Hiervoor hebben we onder andere het probleem van contextspecificiteit geëxploreerd zoals dat eigen is aan expert performance, onderwijs-context en toetsmethoden. We bespreken eerst wat bekend is over betrouwbare en valide toetsing van klinisch redeneren en we introduceren het probleem van contextspecificiteit, dat het meten van klinisch redeneren in de klinische praktijk zo moeilijk maakt. Vervolgens gaan we dieper op het begrip 'context' in en we introduceren twee theorieën die hierbij een centrale rol spelen: situated cognition en ecological psychology. Deze theorieën vormen een centrale denklijn in alle studies in deze dissertatie. Kern in deze twee theoretische visies is dat men nooit vat kan krijgen op het klinisch redeneren zonder de context in acht te nemen, vooral omdat het redeneren zo sterk ingebed is in ervaring (actoren, setting en hun interacties). Belangrijk is dan ook, dat we niet geprobeerd hebben om het klinisch redeneren te meten vanuit een standaard psychometrisch oogpunt – onze nadruk lag juist op het beter begrijpen van dit alomtegenwoordige fenomeen in medisch onderwijs. Zodoende hebben we in dit proefschrift onderzoek gedaan naar de invloed van context op de prestaties bij het klinisch redeneren. In dit licht hebben we het belang van theoretische kaders als lens voor verder onderzoek willen benadrukken.

Hoofdstuk 2 start met de uitdagingen/problemen die artsen dagelijks tegenkomen – het idee dat onderwijs en praktijk plaatsvinden in complexe omgevingen. We bespreken hoe studies uit disciplines andere dan geneeskunde gebruikt kunnen worden om beter grip te krijgen op het concept van context in het klinische consult. Allereerst hebben we gekeken naar het gebruik van de term 'context' in het kader van een klinisch consult in relatie tot medisch onderwijs. Vervolgens zijn we dieper ingegaan op de term 'context' in diverse velden buiten de geneeskunde, zoals wiskunde, natuurkunde en psychologie. We hebben hieruit een definitie gedestilleerd voor medisch onderwijs die we gebruiken als uitgangspunt voor alle studies in dit proefschrift. Tevens introduceren we twee voor ons belangrijke theorieën: situated cognition and ecological psychology, met als doel het bestuderen van medisch onderwijs. De theorieën illustreren hoe individuele systeemelementen, of contextuele factoren (patiënt, arts en consult), gerelateerd kunnen worden aan het systeem in zijn algemeenheid op potentieel complexe manieren. De benadering van contextuele factoren vanuit dit perspectief – patiënt, arts, consult en hun interacties – passen we consistent toe in dit proefschrift. Vervolgens besteden we specifiek aandacht aan deze factoren en hun interacties in de klinische arena. We belichten problemen binnen en implicaties voor het gebied van medisch onderwijs. Een essentieel punt is dat deze theorieën beargumenteren dat we niet-lineaire benaderingen moeten gebruiken als we contextuele factoren willen bestuderen.

Tenslotte suggereren we vervolgstappen voor onderzoek naar contextinvloeden voor dit proefschrift en voor medisch onderwijs in zijn algemeenheid.

Hoofdstuk 3 snijdt een voorbeeld aan van een specifieke artsgerelateerde contextuele factor, namelijk het ouder worden. We onderzoeken deze factor vanuit het standpunt van het raamwerk van situated cognition, met als doel potentiële problemen and implicaties van ons theoretisch model beter te begrijpen bij ouderwordende artsen. We gaan specifiek in op diverse zaken die gerelateerd zijn aan het klinische redeneren en de implicaties voor de ouderwordende arts. Dit betreft ook de mogelijke impact van het ouder worden op expertise, informatieverwerking, neurobiologie, intelligentie en zelfregulatie van het leren. We gaan er vanuit dat het werk dat in deze wetenschapsgebieden is gedaan ons kan helpen bij het beter begrijpen van het klinisch redeneren van de ouderwordende arts, en op die wijze ons wat kan leren over effectievere aanpakken voor nascholing en levenslang leren in de geneeskunde. We sluiten af met implicaties voor het gezondheidszorgsysteem en suggesties voor verder onderzoek.

Hoofdstuk 4 beschrijft een kwantitatieve studie waarin we drie verschillende gezichtspunten van dezelfde situatie (patiënt, arts, opleiding) onderzoeken, om te bepalen of deze verschillende betrokken groepen dezelfde situatie anders zien. Situated cognition en ecologische psychologie stellen dat deze drie groepen verschillende, doch aan elkaar gerelateerde visies zouden moeten hebben, omdat ze verschillende affordances¹ en effectiviteiten² (termen uit de ecologisch psychologie) hebben, en het dus redelijkerwijze te verwachten is dat ze elk op een unieke wijze interageren met patiënt, setting van het consult en elkaar (situated cognition). We hebben een confirmatorische factoranalyse gedaan op een drie-componenten OSCE station om vast te stellen wat de patiënt-, de artsen en de opleidingsvisies waren op dezelfde situatie. In overeenstemming met onze theoretische raamwerken bleken de prestatiebeoordelingen - gemeten met terciëlscores - behoorlijk te variëren tussen de componenten van het OSCE station. We concluderen hieruit dat het verkrijgen van meerdere perspectieven van hetzelfde consult een meer valide maat is van de klinisch prestaties van een student. Deze conclusie wordt ondersteund door onze theoretische raamwerken en de bevindingen in onze studies.

Hoofdstuk 5 is een kwantitatieve studie die bewijs aandraagt voor de uitvoerbaarheid, de betrouwbaarheid en de validiteit van een post-encounter form (PEF) om het klinisch redeneren te toetsen. Dit PEF is ook gebruikt in hoofdstuk 2 en de hierna volgende hoofdstukken. In het dit hoofdstuk 5 wordt de psychometrische onderbouwing voor een vrije-tekst PEF bij een OSCE voor eindtweedejaars medisch studenten bestudeerd. De bruikbaarheid, gemeten aan de tijd die voor het invullen nodig was, was goed. De inter-beoordelaars-betrouwbaarheid, bepaald met Cohen's Kappa, was hoog. Aanwijzingen voor de validiteit

¹ Er is geen goede Nederlandse term voor affordance; een affordance is een kwaliteit van een object of een situatie die het mogelijk maakt om een actie te verrichten. Het is dus geen eigenschap van de gebruiker maar van het object of de situatie

² Effectiviteiten zijn eigenschappen van de gebruiker van een object of een situatie die het mogelijk maken de affordances van het object of de situatie te gebruiken.

werden verkregen door univariate vergelijkingen tussen de individuele items op het PEF en andere eind-tweedejaarstoetsresultaten. We komen tot de conclusie dat de psychometrische kwaliteiten van de PEF voldoende waren en dat we het dus kunnen gebruiken bij de toetsing van klinisch redeneren. Dientengevolge is de PEF gebruikt in de volgende hoofdstukken.

Hoofdstuk 6 beschrijft een mixed-methods studie, waarin de invloed van bepaalde patiënt- en consultgerelateerde contextuele factoren (slecht Engels sprekende patiënt, patiënt die de kwaliteit van de arts in twijfel trekt, incorrecte suggesties ten aanzien van diagnose of therapie en atypische presentatie van de ziekte) onderzocht wordt. In het kwantitatieve deel (hoofdstuk 7 beschrijft het kwalitatieve deel) laten we zien dat het probleem van contextspecificiteit aanwezig is in onze bevindingen. We hebben een gemodificeerde versie van de PEF (zie hoofdstukken 4 en 5) als uitkomstmaat gebruikt. Ook hebben we een kwantitatieve scoringsmethode ontwikkeld en toegepast op de hardopdenkprotocollen. In beide componenten hebben (boardgecertificeerde) expert internisten een serie van video-opnames bekeken, waarin de hoofdklachten van drie veelvoorkomende aandoening in de interne geneeskunde getoond werden (HIV-positieve patiënt, screening op colorectale kanker en diabetes mellitus). Hierbij werden bepaalde patiëntgebonden, artsgerelateerde en consultgerelateerde contextuele factoren gemodificeerd. In aansluiting aan iedere video-presentatie vulden de deelnemers de PEF in en bekeken dan de video een tweede keer waarbij ze een formeel hardopdenkprotocol volgden. Het effect van de specifieke contextuele factoren leek klein tot matig te zijn, met r -waardes van 0,2 tot 0,4. Twee belangrijke bevindingen kwamen in deze studie naar boven, die de idee ondersteunen dat cognitive load bijdraagt aan het probleem van contextspecificiteit. Ten eerste werden negatieve correlaties tussen de aanwezigheid van contextuele factoren en PEF-performance vaker gezien wanneer deelnemers de contextuele factor verbaliseerden gedurende de hardopdenkanalyse. Ten tweede, waren de voorspelde surrogaat-eindpunten voor cognitive load bij het hardop denken (incorrect gebruik van sematic qualifiers, oppervlakkige probleemrepresentaties en pauzes) geassocieerd met lagere niveaus van klinisch redeneren zoals gemeten met de PEF.

Hoofdstuk 7 is de kwalitatieve component van de studie die in hoofdstuk 6 beschreven is. Het doel van deze studie was te onderzoeken hoe contextuele factoren klinisch redeneren beïnvloeden. De kwalitatieve data van de experts werden verzameld uit twee bronnen: hardopdenkprotocollen en vrije-tekstvragen waarin de (boardgecertificeerde) internisten gevraagd werden om commentaar te geven op de invloed van de gemanipuleerde contextuele factoren op hun klinische redeneerprocessen. De databronnen leverden zowel daadwerkelijke performance-informatie (hardopdenkprotocollen) als opinies en reflecties ten aanzien van de invloed van context op klinisch redeneren (vrije-tekstantwoorden). Met gebruikmaking van een “grounded theory” aanpak werden de resultaten van beide databronnen geanalyseerd op opkomende thema’s. Deze thema’s werden vervolgens gecombineerd tot een uniform theoretisch model. Diverse thema’s kwamen bovendien, ze zijn grofweg te classificeren als *componenten* die de impact van contextuele factoren beïnvloeden, *mechanismen* om met deze factoren om te gaan en *consequenties* van de contextuele factoren op de patiëntenzorg. Thema’s van beide databronnen overlaptten goed met elkaar, wat suggereert dat experts op zijn minst voor een deel op de hoogte zijn van mogelijke in-

vloeden van context op hun redeneerprocessen. Opmerkelijke uitzonderingen waren de thema's 'missed key findings', 'balancing goals' en de 'influence of encounter setting', die wel in de hardopdenkprotocollen gezien werden, maar niet in de vrije-tekstantwoorden. 'Missed key findings' traden alleen op als deelnemers één of meer contextuele factoren verbaliseerden in de hardopdenkprotocollen, wat suggereert dat cognitive load een rol speelt. Ons uniform model is in overeenstemming met de grondgedachten van cognitive load, situated cognition en ecological psychology. Een groot aantal mogelijk manipuleerbare invloeden van contextuele factoren op klinisch redeneren worden besproken. De resultaten van deze studie suggereren dat een belangrijke component van contextspecificiteit is dat contextuele factoren (patiënt, consult, arts), anders dan key-features voor het stellen van de diagnose, een grote invloed hebben op de kwaliteit van het redeneerproces.

Hoofdstuk 8 beschrijft een deel van een tweedelige studie (hoofdstukken 8 en 9) bestaand uit een gerandomiseerde cross-over trial, waarin de invloed van authenticiteit van de leeromgeving op klinische redeneeruitkomsten van tweedejaarsstudenten geneeskunde wordt bestudeerd. We hebben de balans tussen toenemende authenticiteit en cognitieve load onderzocht. In essentie hebben we het probleem van contextspecificiteit vanuit het standpunt van instructie bekeken; is de onderwijsvorm van invloed in de zin dat het leidt tot variabiliteit in de prestaties van studenten tussen verschillende leeromgevingen? Er werd geen effect van toenemende authenticiteit van de leeromgeving (papieren casus, DVD-casus of simulatiepatiënt presentatie van dezelfde inhoud) gevonden; geen van de uitkomstmaten bereikte voldoende statistische significantie. Deze bevindingen suggereren dat de balans tussen toenemende authenticiteit van instructie en het reduceren van cognitive load complexer is dan eerst gedacht, met name in de preklinische setting. Medisch onderwijskundigen moeten daarom zich goed bewust zijn van de noodzaak om bij het ontwerpen van nieuwe onderwijsvormen een goed evenwicht te vinden tussen toenemende authenticiteit en factoren zoals cognitive load, leerstof en ervaringen van de lerende. Een andere belangrijke implicatie is dat, ondanks de intense marketing van de simulatie-industrie en onze wens om in toenemende mate onze specialisten in opleiding technologiegeoriënteerde te laten zijn, het simpelweg verhogen van de authenticiteit zeker geen panacee genoemd mag worden.

Hoofdstuk 9 onderzoekt hetzelfde cohort medische studenten als in hoofdstuk 8 maar nu in hun derde jaar van de medische opleiding. De onderwijsvorm die ze in het tweede jaar genoten hebben (zie hoofdstuk 8) had geen invloed op hun klinisch redeneren in het derde jaar. Wat van invloed was, was het aantal patiënten dat ze gezien hadden gedurende hun derdejaars co-schap. Een opvallende bevinding was dat er geen cross-overeffect was ten aanzien van onze uitkomstmaten – het zien van meer patiënten met anemie was niet geassocieerd met betere performance op diabetes- of buikpijnpatiënten. Dit hoofdstuk draagt bij aan de literatuur doordat het aantoont dat authenticiteit in preklinisch onderwijs geen invloed heeft op het co-schapjaar. Echter, de patiënten die in de praktijk gezien zijn (wat beschouwd mag worden als de authentiekste vorm van alle) heeft wel een invloed op de uitkomsten van redeneren gedurende de co-schappen.

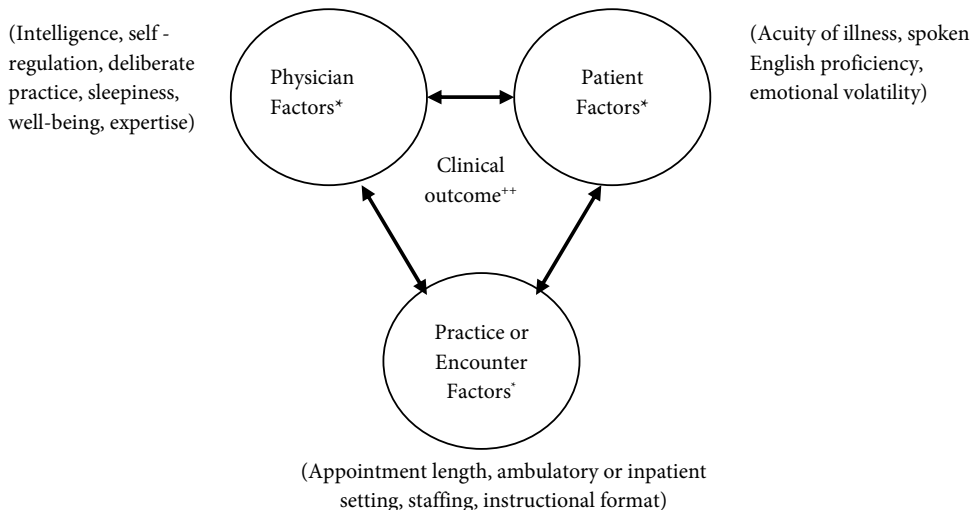
Hoofdstuk 10 bestudeert de problemen die contextspecificiteit stelt aan experts, waarbij we gebruikmaakten van een nieuwe meettechniek, namelijk functionele magnetic resonance imaging (fMRI). In essentie is de onderliggende aanname hierbij, dat een deel van het contextspecificiteitsprobleem zou kunnen liggen in het feit dat onze maten onvoldoende precies zijn om de interne mentale processen te vangen. In dit hoofdstuk hebben we fMRI en hardopdenkprotocollen gecombineerd om te bepalen of functionele neuro-imaging correleert met analytische en non-analytische theorieën over redeneren - analoog aan het concept van automatisering in cognitive load theorie. Ook hebben we deze combinatie gemaakt om te bepalen of er functionele neuro-imagingverschillen zijn het beantwoorden en het hardopdenken over vignettegebaseerde toetsvragen. Boardgecertificeerde internisten beantwoordden 'live' in een fMRI-scanner vragen uit een internistenexamen, gevolgd door een hardopdenkprocedure. Bewijs, in de vorm van veranderingen op de fMRI, werd gevonden voor de aanwezigheid van zowel analytische als non-analytische redeneersystemen. We hebben ook aanwijzingen gevonden voor de zinvolheid van het gebruik van hardopdenkprotocollen om de interne mentale processen - zoals klinisch redeneren - in kaart te brengen. Dit model om fMRI te gebruiken kan niet alleen ons begrip van de theorie versterken, maar ook onze praktijk van doceren en toetsen van klinisch redeneren.

Hoofdstuk 11 verklaart hoe alle voorgaande hoofdstukken bijdragen aan een verder begrip van de invloed die contextuele factoren hebben op klinisch redeneren gedurende het klinische consult. In dit hoofdstuk gaan we terug op onze originele centrale onderzoeksvraag, namelijk of contextuele factoren een invloed hebben op de klinisch redeneren performance zowel bij experts als bij specialisten in opleiding. Het antwoord dat ons werk levert is, dat deze contextuele factoren een invloed hebben op het klinisch redeneren, zowel bij de expert als bij de lerende. 'Context' in het klinische consult is een complex systeem. Het heeft zowel sociale (actoren) als fysieke (setting) constructen, die alle ook nog evolueren in tijd. Deze dissertatie leidt tot een beter theoretisch begrip, maar ook tot empirische bevindingen, om grip te krijgen op dit fenomeen. De hoofdfocus van dit proefschrift was het verbeteren van ons begrip van het probleem 'contextspecificiteit' in relatie tot klinisch redeneren, en diverse puzzelstukjes zijn besproken. We hebben onder andere meer kennis over contextspecificiteit eigen aan expert performance, onderwijscontext en toetsmethodieken. Het heeft achtereenvolgens de twee overrijdende theoretische raamwerken beschreven, die bijgedragen hebben aan ons begrip van contextuele factoren en hoe cognitive load theory naar boven kwam als een belangrijk derde raamwerk. Vervolgens worden de conclusies besproken en de implicaties, beperkingen en aanbevelingen voor verder onderzoek. Samenvattend waren onze conclusies in overeenstemming met voorspelling volgens de drie theoretische kaders and diverse gebieden voor verder onderzoek worden besproken.

Appendices

Appendix 1:

Context in the Clinical Encounter and Proposed Results on the Clinical Outcome (Clinical Reasoning): A Situated Cognition and Ecological Psychology Perspective[^]



^{*}=Examples of each factor (physician, patient, and practice) are shown next to each factor. Practice factors and encounter factors are used interchangeably; both refer elements pertaining to the environment or setting for the encounter.

^{*}=factors that interact as shown by arrows; parentheses next to each factor are examples or factor elements

⁺⁺=clinical outcome is dependent upon these 3 factors, their interactions, and possibly other inputs.

[^]=Situated cognition takes the approach of the individual and the environment and all of the above interactions can and do influence the outcome (patient care) in such a model; individual participants are influenced by affordances, effectivities and goal structures/hierarchy (ecological psychology)

Appendix 2:

New Onset Diabetes Mellitus OSCE Case

Patient Complaint(s): Fatigue, Polyuria, Thirst, Weight Loss

Actual Diagnosis: Adult Onset Diabetes Mellitus

Patient Presentation: Concerned, quiet but cooperative middle-aged woman/ man in office visit. He/she is seated in the exam room with a bottle of water, and should be intermittently drinking throughout the interview.

Opening Statement by patient: "I have been unusually thirsty and tired lately."

Appropriate examinee level as written: Examinee is MS2 OSCE student

Necessary equipment, props, makeup, costume, etc.: **SP should wear socks during exam, bottle of drinking water.**

Additional materials required such as lab results, x-rays, scans, etc.: None

Case Synopsis: A 45-55 year old man presents to the acute care clinic with a complaint of fatigue and thirst.

Examinee objective: To identify the classic presentation of new onset diabetes, and develop an understanding of the key historical and physical exam findings necessary to distinguish this presentation from other etiologies of fatigue, polyuria, and polydipsia. To identify common complications of diabetes mellitus through history and physical exam findings.

Standardized patient recruiting requirements: Actor should be moderately overweight male between the age of 45 and 55 of any race, without recent surgical scars

Adaptable to other sexes? YES, prefer male Other races? YES Other ages? YES

In a new project or the same project? New project.

Standardized Patient Training Summary

Opening statement: “I’ve been unusually thirsty and tired lately.” (take a swig from your water bottle as student opens door)

Follow-up: “I’ve been waking up a lot.”

Make sure your shoes are off and your socks are on.

Current Complaint:

You are Eliza/Elwood Taylor, a 51 year old elementary school teacher seeing a physician because of increased thirst for the past month. You also report feeling more fatigued over this same time frame, but have attributed this to you sleep being interrupted by getting up to go to the bathroom several times nightly. If asked to quantify the number of time you are going to the bathroom, you state you are urinating more frequently than usual (7-8 times a day) and getting up 3-4 times a night to urinate. You are able to get back to sleep, but are sure that the broken sleep is contributing to your tiredness. If asked, you deny any weakness in your urine stream, and always feel that you completely empty your bladder. You have also noticed that the volume of each urination is greater than usual, but deny burning with urination if asked. You are excessively thirsty, drinking a big glass of water every half hour or so. You’ve tried to drink less, and have even cut out coffee and alcohol, hoping to urinate less often, but the frequent urination persists and you just get thirstier. If asked, you deny any weight gain or loss, but do note an increase in your appetite. You’ve also noted some difficulty reading due to blurry vision, but thought this was probably just eye strain related to your job and you plan to see an optometrist. When given the eye chart test, the smallest type you can see is 20/50. Finally, you’ve noted some intermittent tingling sensations in your feet over the last few weeks.

Social History:

You are a third grade teacher in the Montgomery County Public Schools (Rock Creek Forest Elementary School on Grubb Road). You have a class of 28 students with a ½ time reading aide. You enjoy your work and love the children.

You are married. Your husband/wife/partner of 20 years is Pat. You have a good relationship, but over the last month you’ve lost interest in sex because of erectile dysfunction (males) or pain due to vaginal dryness (females) and you are worried about this.

You have two children, who are healthy and well. You are finding lately that it’s hard to climb a few steps from the parking lot to the school. At recess time, you would rather just stay at your desk than scurry down to the lounge to duplicate papers like you usually like to do. You don’t exercise as much; but you do walk around the block with your husband 3 or 4 times a week in the evenings when your knee isn’t acting up.

You usually drink one cup of coffee in the morning and rarely drink any soda. The composition of your diet has not changed recently, but you’ve been hungry and eating more than usual.

You have never been out of the country and haven't been around anyone else who is sick except for all of your runny-nose-and-coughing students.

Past Medical History:

You see a doctor at the family health clinic every year for a checkup and TB test as required by the county for teachers. Your last checkup was about 9 months ago and you were told that everything was fine. You have had "borderline" high blood pressure and high cholesterol (over 250) in the past, but have not required any medications. You keep up with all your routine health care, and have not had any issues come up. You have had no surgeries recently (can work in surgical scars or c-section scars as hysterectomy). If female, you are still menstruating and don't know how old your mother was when she entered menopause.

Family History:

Only if asked about a family history of diseases do you reveal that your mother is now 78 and has congestive heart failure, your father is deceased, having died at 55 of a massive heart attack, and you have an older sister, Peggy (age 57) who was diagnosed with diabetes when she was your age. You don't know what kind of diabetes she has nor which medication she takes. You know of no other illnesses that run in your family.

Timeline

Last several weeks (3 or 4) – You have noticed the very gradual onset of increased fatigue, thirst, and urination both during the day and at night. You admit these symptoms may have been going on for longer, but you have just noticed them being more pronounced over this past month.

Physical Examination Findings

If asked: when you checked in, you felt a little lightheaded when the nurse stood you up to take your blood pressure.

When seated: you should appear mildly anxious, and drink from your water bottle intermittently throughout the interview.

If asked to perform the eye test with a card: you will only be able to read down to the 20/50 line in both eyes with your corrective glasses (student should allow you to use your glasses if needed).

If the student tests your sensation to light touch, you are unable to feel light touch on the bottoms of your feet bilaterally. If the student uses a tuning fork to assess sensation to vibration, this is normal.

125Appendix 3: Standardized Patient (SP) Checklist[^]*

Opening statement: "I've been feeling especially tired lately."

History	Yes	No
Upon questioning by the student, I responded that:	(1 point each)	
1. (CS) I'm having trouble with my job because I feel tired or worn out "Describe your tiredness."	O	O
2. (CS) I'm so thirsty I'm drinking a water every half hour or so. "Have you been excessively thirsty?" "Are you drinking more than usual?"	O	O
3. (CS) I have been urinating more frequently than usual (7 or 8 times a day). I am waking up 3 or 4 times at night, but only because I need to urinate. "Are you going to the bathroom more often?" "Has your urination changed lately?"	O	O
4. (PN) In the last month, my weight has not changed. "Have you had any weight change lately?" "Have you lost any weight lately?"	O	O
5. (AS) I've had some difficulty reading due to blurry vision.	O	O
7. (AS) I've noticed some numbness in my feet.	O	O
8. (AS) I've lost interest in sex because of erectile dysfunction (male) or dryness (female). "How has this affected your life?" "What additional concerns do you have?"	O	O
9. (RF) My older sister was diagnosed with diabetes when she was my age. "Does anyone in your family have diabetes?"	O	O
10. (RF) I have had "borderline" high blood pressure OR high cholesterol in the past (either is acceptable)	O	O
11. (PN) I don't have pain or burning when I urinate.	O	O
Physical Examination		
<i>The student:</i>		
13. Washed his/her hands in my presence before examining me or putting on gloves.	O	O (1 pt)
14. Vision: The student testing my vision by		
a. Testing each eye separately using a vision card	O	O (1 pt)
b. Checked my vision, but tested both eyes simultaneously or without a card	O	O (0 pt)
15. Eye examination: The student examined my eyes		
a. Using an ophthalmoscope	O	O (1 pt)
b. Examined my eyes, but did not use an ophthalmoscope	O	O (0 pt)
16. Feet: The student examined my feet:		
a. Both feet while barefoot	O	O (1 pt)
b. Only one foot, or with socks on	O	O (0 pt)
17. Sensation: The student tested the sensation in my feet by:		
a. Using a tuning fork	O	O (1 pt)
b. Using a cotton swab or other object	O	O (1 pt)
c. Both	O	O (2 pt)

*: CS=cardinal symptoms, AS=associated symptoms, PN=pertinent negatives, RF=risk factors

[^]=items 6 and 12 had been eliminated from use prior to the study without renumbering the form

COMMUNICATION & INTERPERSONAL SKILL RATING FORM

New Communications and Interpersonal Skills Checklist.
Based on ABIM "Humanistic Qualities".

Rating scale: 1-5	(1=poor, 2=fair, 3=good, 4=very good, 5=excellent)				
1. The student introduced themselves properly, warmly and addressed you by your surname.	1	2	3	4	5
2. The student treated you with respect as one would want to be treated themselves.	1	2	3	4	5
3. The student used appropriate open-ended and thoughtful questions.	1	2	3	4	5
4. The student listened intently, let you tell your story and did not interrupt.	1	2	3	4	5
5. The student showed interest in your ideas, values and concerns.	1	2	3	4	5
6. The student used words you could understand and explained terms in plain language.	1	2	3	4	5
7. How would you overall rate this student's communication skills?	1	2	3	4	5

Appendix 4: Post Encounter Form

Summary statement (1 point. In one sentence, summarize the most important historical and physical exam features in this patient's presentation.)

Problem List (3 points, one point for each reasonable problem listed)

- 1.
- 2.
- 3.

Differential Diagnosis (3 points; list diagnoses that would explain **all** the problems you've listed)

- 1.
- 2.
- 3.

MOST likely Diagnosis (1 point)

- 1.

Supporting Data for MOST likely Diagnosis (2 points)

Provide 4 key history and/or PE facts to support the MOST likely diagnosis. One point for each fact provided.

- 1.
- 2.
- 3.
- 4.

Appendix 5: Faculty Rating Form for Oral Presentation Station

I. Oral Presentation Skills

	Unacceptable	Needs Improvement	Acceptable	Above Average	Outstanding
Pace	<input type="checkbox"/> - Does not complete	<input type="checkbox"/> - Less than 5 minutes and incomplete - More than 7 Minutes; extraneous data	<input type="checkbox"/> - 5-7 Minutes - Less than five minutes, but complete - More than seven minutes, but minimal extraneous data		
HPI	<input type="checkbox"/> - Chronology is poor/chaotic - Poor characterization of the symptom - No pertinent negatives	<input type="checkbox"/> - Inadequately describes chronology - Superficial characterization of symptoms - No pertinent negatives	<input type="checkbox"/> - Adequately describes chronology of the illness - Adequate characterization of CC - Ends with a few pertinent negatives	<input type="checkbox"/> - Chronology is clearly stated - Full characterization of the CC - Demonstrates <u>recognition</u> of important details	<input type="checkbox"/> - Chronology is clearly stated - Full characterization of the CC - Demonstrates <u>prioritization</u> of important details
Context: SH/PMH/PSH	<input type="checkbox"/> - Major omissions - Completely disorganized	<input type="checkbox"/> - Superficial; lacking important basic details - Poorly organized	<input type="checkbox"/> - Generally maintains format - Few or minor omissions	<input type="checkbox"/> - Maintains format - Complete; no omissions - Demonstrates recognition of important details	<input type="checkbox"/> - Maintains format - Complete; no omissions - Demonstrates prioritization of important details
Physical Exam	<input type="checkbox"/> - Major omissions - Completely disorganized	<input type="checkbox"/> - Superficial; lacking important basic details - Poorly organized	<input type="checkbox"/> - Generally maintains format - Adequate characterization of presenting symptom - Describes all Hx and PE elements in adequate detail - Occasional omissions /errors	<input type="checkbox"/> - Maintains format - Complete; no omissions - Demonstrates recognition of important details	<input type="checkbox"/> - Maintains format - Complete; no omissions - Demonstrates prioritization of important details

Overall Organization	<input type="checkbox"/> <ul style="list-style-type: none"> - Completely disorganized; rambling - Minimal or no attempt to follow prescribed format - Major omissions 	<input type="checkbox"/> <ul style="list-style-type: none"> - Disorganized, but attempts to follow the basic format - Superficial; lacking important basic details - Unduly repetitive 	<input type="checkbox"/> <ul style="list-style-type: none"> - Follows basic format - Most elements well organized - Few or minor omissions - Minor repetition 	<input type="checkbox"/> <ul style="list-style-type: none"> - Smooth presentation - Minimal or no repetition - Demonstrates ability to recognize important data 	<input type="checkbox"/> <ul style="list-style-type: none"> - Fluent presentation - Minimal use of notes - Demonstrates ability to prioritize data
-----------------------------	--	---	---	--	---

1. **Summary statement** (quality of a 1 to 3 sentence summary of patient’s reason for presentation)

Unable to summarize Poor/inadequate summary Adequate summary Well summarized; recognizes key details Outstanding summary; demonstrates understanding

2. **Semantic Competence** (i.e. use of terminology to efficiently and concisely convey patient’s symptoms and findings)

Uses lay terms patient’s words Incorrect use of medical language Correctly uses some medical terminology Frequently and correctly uses medical terminology Advanced fluency; Eloquent, concise

Dedications

To my wife, Kristen, and two sons, Andrew and Daniel, for their unwavering love and support and for sacrificing so much time in my pursuit of this degree.

To my mother and Richard, for believing in me, advising me and encouraging me to always follow my dreams.

To my father for instructing me to focus my time and efforts toward this degree.

To my sister for her advice, friendship, and support.

To my in-laws and Walt for their prayers and discernment.

To my promoters and Tony Artino for their direction and encouragement so that I could complete this thesis.

Proverbs 3:5-6.

Curriculum Vitae

Steven J. Durning received his college degree from The Pennsylvania State University in 1991. He then went to medical school at the University of Pittsburgh and received his MD degree in 1995. Next, he completed his residency training in Internal Medicine at Wright-Patterson Air Force Base in Dayton, Ohio in 1998. After completing residency, he served as the chair of the education committee for the internal medicine residency program and as the site clerkship director for the Uniformed Services University (USU). In 2002, Dr Durning moved to Maryland to serve as a second-year medical student course director at the Uniformed Services University in Bethesda, Maryland. While serving as the director of the Introduction to Clinical Reasoning course, he started his PhD project in 2007 addressing the influence of contextual factors on clinical reasoning. In addition to serving as a course director, he is the principal investigator of the Long Term Career Outcome Study at USU. Dr Durning is a professor of medicine and pathology and he serves on a number of national and international organizations. He is the principal or associate investigator on a number of grants, and has published over 100 peer-reviewed manuscripts.

SHE dissertation series

In the SHE Dissertation Series dissertations are published of PhD candidates from the School of Health Professions Education (SHE) who defended their PhD thesis at Maastricht University. The most recent ones are listed below. For more information go to: www.maastrichtuniversity.nl/she.

- De Leng, B. (8-12-2009). Wired for learning. How computers can support interaction in small group learning in higher education.
- Maiorova, T. (29-05-2009). The role of gender in medical specialty choice and general practice preferences.
- Bokken, L. (04-03-2009). Innovative use of simulated patients for educational purposes.
- Wagenaar, A. (18-09-2008). Learning in internships. What and how students learn from experience.
- Driessen, E. (25-06-2008). Educating the self-critical doctor. Using portfolio to stimulate and assess medical students' reflection.
- Derkx, H. (18-06-2008). For your ears only. Quality of telephone triage at out-of-hours centres in the Netherlands.
- Niessen, Th. (30-11-2007). Emerging epistemologies: making sense of teaching practice.
- Budé, L. (05-10-2007). On the improvement of students' conceptual understanding in statistics education.
- Niemantsverdriet, S. (26-07-2007). Learning from international internships: A reconstruction in the medical domain.
- Marambe, K. (20-06-2007). Patterns of student learning in medical education – A Sri Lankan study in traditional curriculum.
- Pleijers, A. (19-01-2007). Tutorial group discussion in problem-based learning.
- Sargeant, J. (21-09-2006). Multi-source feedback for physician learning and change.
- Dornan, T. (12-06-2006). Experience-based learning.
- Wass, V. (12-05-2006). The assessment of clinical competence in high stakes examinations.
- Prince, K. (21-04-2006). Problem-based learning as a preparation for professional practice.

ICO dissertation series

In the ICO Dissertation Series dissertations are published of graduate students from faculties and institutes on educational research within the following universities: Eindhoven University of Technology, Leiden University, Maastricht University, Open University of the Netherlands, University of Amsterdam, University of Groningen, University of Twente, Utrecht University, VU University Amsterdam, and Wageningen University (and formerly Radboud University Nijmegen and Tilburg University). Below is a list of the most recent dissertations For more information go to: www.ou.nl/ico.

179. Munneke-de Vries, E.L. (11-01-2008). Arguing to learn: Supporting interactive argumentation through computer-supported collaborative learning. Utrecht: Utrecht University.
180. Nijveldt, M.J. (16-01-2008). Validity in teacher assessment. An exploration of the judgement processes of assessors. Leiden: Leiden University.
181. Jonker, H.G. (14-02-2008). Concrete elaboration during knowledge acquisition. Amsterdam: VU University Amsterdam.
182. Schuitema, J.A. (14-02-2008). Talking about values. A dialogue approach to citizenship education as an integral part of history classes. Amsterdam: University of Amsterdam.
183. Janssen, J.J.H.M. (14-03-2008). Using visualizations to support collaboration and coordination during computer-supported collaborative learning. Utrecht: Utrecht University.
184. Honingh, M.E. (17-04-2008). Beroepsonderwijs tussen publiek en privaat: Een studie naar opvattingen en gedrag van docenten en middenmanagers in bekostigde en niet-bekostigde onderwijsinstellingen in het middelbaar beroepsonderwijs. Amsterdam: University of Amsterdam.

185. Baartman, L.K.J. (24-04-2008). Assessing the assessment: Development and use of quality criteria for competence assessment programmes. Utrecht: Utrecht University.
186. Corbalan Perez, G. (25-04-2008). Shared control over task selection: Helping students to select their own learning tasks. Heerlen: Open University of the Netherlands.
187. Hendrikse, H.P. (22-05-2008). Wiskundig actief: Het ondersteunen van onderzoekend leren in het wiskunde onderwijs. Enschede: University of Twente.
188. Moonen, M.L.I. (26-09-2008). Testing the multi-feature hypothesis: Tasks, mental actions and second language acquisition. Utrecht: Utrecht University.
189. Hooreman, R.W. (18-11-2008). Synchronous coaching of the trainee teacher: An experimental approach. Eindhoven: Eindhoven University of Technology.
190. Bakker, M.E.J. (02-12-2008). Design and evaluation of video portfolios: Reliability, generalizability, and validity of an authentic performance assessment for teachers. Leiden: Leiden University.
191. Kicken, W. (12-12-2008). Portfolio use in vocational education: Helping students to direct their learning. Heerlen: Open University of the Netherlands.
192. Kollöffel, B.J. (18-12-2008). Getting the picture: The role of external representations in simulation-based inquiry learning. Enschede: University of Twente.
193. Walraven, A. (19-12-2008). Becoming a critical websearcher: Effects of instruction to foster transfer. Heerlen: Open University of the Netherlands.
194. Radstake, H. (14-05-2009). Teaching in diversity: Teachers and pupils about tense situations in ethnically heterogeneous classes. Amsterdam: University of Amsterdam.
195. Du Chatenier, E. (09-09-2009). Open innovation competence: Towards a competence profile for inter-organizational collaboration in innovation teams. Wageningen: Wageningen University.
196. Van Borkulo, S.P. (26-06-2009). The assessment of learning outcomes of computer modelling in secondary science education. Enschede: University of Twente.
197. Handelzalts, A. (17-09-2009). Collaborative curriculum development in teacher design teams. Enschede: University of Twente.
198. Nievelstein, F.E.R.M. (18-09-2009). Learning law: Expertise differences and the effect of instructional support. Heerlen: Open University of the Netherlands.
199. Visser-Wijnveen, G.J. (23-09-2009). The research-teaching nexus in the humanities: Variations among academics. Leiden: Leiden University.
200. Van der Rijst, R.M. (23-09-2009). The research-teaching nexus in the sciences: Scientific research dispositions and teaching practice. Leiden: Leiden University.
201. Mainhard, M.T. (25-09-2009). Time consistency in teacher-class relationships. Utrecht: Utrecht University.
202. Van Ewijk, R. (20-10-2009). Empirical essays on education and health. Amsterdam: University of Amsterdam.
203. Seezink, A. (18-11-2009). Continuing teacher development for competence-based teaching. Tilburg: Tilburg University.
204. Rohaan, E.J. (09-12-2009). Testing teacher knowledge for technology teaching in primary schools. Eindhoven: Eindhoven University of Technology.
205. Kirschner, F.C. (11-12-2009). United brains for complex learning. Heerlen: Open University of the Netherlands.
206. Wetzels, S.A.J. (18-12-2009). Individualized strategies for prior knowledge activation. Heerlen: Open University of the Netherlands.